AD_____

Award Number: W81XWH-04-1-0633

TITLE: Targeting Breast Cancer Cells for Destruction

PRINCIPAL INVESTIGATOR: Jun Ma, Ph.D.

CONTRACTING ORGANIZATION: Cincinnati Children's Hospital Medical Center
Cincinnati, OH 45229

REPORT DATE: July 2005

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
01-07-2005	Annual	1 Jul 2004 – 30 Jun 2005

4. TITLE AND SUBTITLE

Targeting Breast Cancer Cells for Destruction

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-04-1-0633

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Jun Ma, Ph.D.

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Cincinnati Children's Hospital Medical
Center
Cincinnati, OH 45229

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES
Original contains colored plates: ALL DTIC reproductions will be in black and white

14. **ABSTRACT**
The long-term objective of our work is to understand the actions of existing anti-cancer drugs and to investigate concepts leading to the design of new drugs to eradicate breast cancer. One of our long-standing research interests is in understanding the molecular principles that govern biological specificities. We have been studying both protein-protein interactions and protein-DNA interactions through a battery of different approaches including molecular, genetic, biochemical and structural approaches. The analysis of molecular principles governing biological specificities is not only important as a basic science problem but also has implications in human health and disease prevention. We have recently investigated the roles of the molecular chaperone Hsp90 in normal development in Drosophila. Hsp90 is a specific target of the anti-cancer druggeldanamycin. Our experiments have identified a new Hsp90 client protein that is critical for normal embryonic development, thus establishing a foundation for further investigation of the actions of Hsp90 and the anti-cancer drug geldanamycin. We have also proposed to investigate cellular delivery methods for specifically targeting cancer cells. Our design takes advantage of biological specificities and uses a novel double-targeting system to increase cancer cell specificity.

15. SUBJECT TERMS
Targeted therapies, apoptosis, specificity, interaction

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
					USAMRMC
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER *(include area code)*
U	U	U	UU	54	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

Table of Contents

INTRODUCTION

One of our long-standing research interests is in understanding the molecular principles that govern biological specificities. We have been studying both protein-protein interactions and protein-DNA interactions through a combination of different approaches including molecular, genetic, biochemical and structural approaches. In particular, we have been investigating the molecular functions and specificities of homeodomain proteins (human PITX2 and *Drosophila* Bicoid) that are required for normal embryonic development. The analysis of molecular principles governing biological specificities is not only important as a basic science problem but also has implications in cancer prevention. For example, the anti-cancer drug geldanamycin is a specific inhibitor of the molecular chaperone Hsp90 [2]. Understanding how Hsp90 regulates the activities of its client proteins in controlling normal development and cellular physiology represents both a basic biological problem and a medical interest in search for improved therapeutic uses of existing anti-cancer drugs such as geldanamycin. Our recent studies have revealed a new Hsp90 client protein (Bicoid) that is involved in normal development, establishing a foundation for further analysis of the actions of the molecular chaperone Hsp90 and the anti-cancer drug geldanamycin. Moreover, the knowledge in biological specificities can aid the development of new anti-cancer drugs. There are two major challenges to the design of effective drugs against breast and other cancers. First, such drugs must have a high specificity for cancer cells; this is important for killing cancer cells while causing relatively little harm to normal cells. Second, the mechanisms of cancer cell destruction must be effective and designed to minimize ways that cancer cells can develop to escape the drugs: breast cancer cells tend to become resistant to anti-hormone drugs after such treatments. We proposed to investigate cellular delivery methods for specifically targeting cancer cell destruction using a novel double-targeting system. The long-term objective of our work is to understand the actions of existing anti-cancer drugs and to investigate concepts leading to the development of new drugs with high specificity and efficacy to eradicate breast cancer.

BODY

We have made progress in understanding molecular specificities in protein-protein and protein-DNA interactions. These studies reflect our long-standing research interests and are further detailed in the attached publications. Briefly, the research paper by Fu and Ma (2005) describes our analysis of the *Drosophila* transcription activator protein Bicoid (Bcd) and its interaction with the co-factor dCBP. Our results show that dCBP can modulate the activity of Bcd and facilitate the switch between the active and inactive states of Bcd in activating transcription. The publications by Chaney et al. (2005) and Baird-Titus et al. (2006) describe our collaborative studies to determine the solution structures of protein-DNA complexes and to understand the specificity codes in protein-DNA interactions. Our studies reveal that the homeodomains of both human PITX2 and *Drosophila* Bcd exhibit novel structural properties. In addition, our results reveal that the protein-DNA interface responsible for determining the DNA

binding specificity is highly dynamic in both cases. A recently published review article by the PI on transcription on-off switches is also included (Ma, 2005).

As part of our long-term objective toward understanding and improving existing anti-cancer drugs, we recently investigated the roles of the molecular chaperone Hsp90 in normal development in *Drosophila*. Hsp90 is a specific target of the anti-cancer drug geldanamycin [2]. In *Drosophila*, Hsp90 is encoded by the gene *Hsp83* [3]. Our recent unpublished experiments have revealed an interaction between Bcd and Hsp90, suggesting that Bcd is a new client protein of Hsp90. In early *Drosophila* embryos, the maternally-contributed Bcd protein is distributed as an anterior-to-posterior gradient that is responsible activating its target genes such as *hunchback* (*hb*) [4]. As reported previously [5] and confirmed by our data (Fig. 1A), the Bcd-dependent expression of *hb* in the anterior half of the embryo is highly precise, with a standard deviation of its expression border of 1.6%. However, in embryos with a reduced maternal contribution of the *Hsp83* activity, the *hb* expression profile exhibits an increased variability with a standard deviation of between 2.3 and 2.5% (Fig. 1B, C). Furthermore, the antimophic allele *e6D* of *Hsp83* causes the *hb* expression boundary to be expanded toward the posterior, from 48.2% to 52.0% in egg length (Fig. 1B). These results suggest that Hsp90 plays a role in regulating the activity of Bcd *in vivo*.

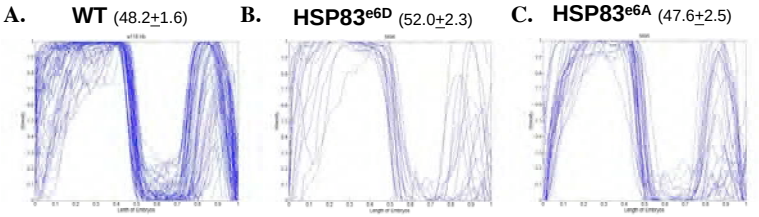

Fig. 1. Increased variability in *hb* expression caused by *hsp83* mutations. Shown are *hb* expression profiles in embryos from wt (A) or $hsp83^{+/-}$ females (B, C). Two different alleles of *hsp83* used are indicated. In our experiments, *hb* expression was detected in embryos by *in situ* hybridization, embryo images captured digitally, staining intensity scanned and plotted. In these plots, the X-axis is the egg length (0-1.0), whereas the Y-axis is the staining intensity (normalized for each embryo to have values between 0-1.0). Each curve in the plots represents the scanned profile of one individual embryo. The mean value of the *hb* expression boundary (embryonic location with 1/2 maximal *hb* expression) and the standard deviation are listed at the top of each plot.

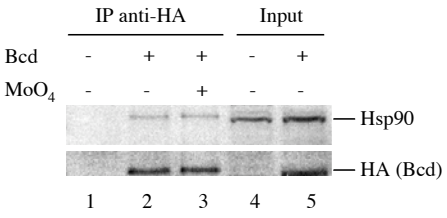

Fig. 2. Co-IP experiments detecting Bcd-Hsp90 interaction. Shown are the results of co-IP experiments showing that Hsp90 is co-precipitated by an antibody (HA) that precipitates a tagged wt Bcd protein in the nuclear extracts of S2 cells expressing the Bcd protein. Western blot shown here was detected by a monoclonal anti-Hsp90 antibody (top) or anti-HA (bottom) to detect the tagged wt Bcd protein. Molybdate (MoO_4) was also included in the co-IP experiment for lane 3. Input represents 10% of the extracts used in co-IP.

To determine whether Bcd and Hsp90 can interact with each other physically, we carried out co-IP experiments in *Drosophila* S2 cells. As shown in Fig. 2, Hsp90 is specifically co-precipitated by an antibody against the HA tag that is attached to the wild type Bcd protein (lane 2). In the absence of Bcd, no Hsp90 is pulled down (lane 1). We have also analyzed the effect of molybdate, another Hsp90 inhibitor that can "lock" or "freeze" the interactions between Hsp90 and some of its client proteins [6]. Our results show that molybdate does not affect the Bcd-

Hsp90 interaction (lane 3), indicating that this is a high affinity binding. These new findings provide the first demonstration that Hsp90 and Bcd can interact with each other. They also establish a foundation for further investigating, with the powerful genetic tools available, the molecular actions of Hsp90 and its inhibitors including the anti-cancer drug geldanamycin.

We proposed to design methods to deliver into cells proteins that can cause cell destruction. The design utilizes the properties of anthrax toxin. This toxin contains three components: two enzymatic proteins lethal factor (LF) and edema factor (EF), and the protective antigen (PA), which is responsible for translocating both LF and EF into the cytosol [7]. Such a translocation activity of PA is strictly dependent on its cleavage by cell-surface furin or furin-like proteases [7]. A recent study has shown that the replacement of the furin cleavage site with a urokinase plasminogen activator (uPA) target site can alter the cleavage/activation specificity of PA [8]. Such an engineered PA is cleaved/activated by uPA, which, along with its receptor (uPAR), is highly expressed in malignant cells [9]. Our design is to use such an altered-specificity PA of anthrax toxin to translocate cell-destructing proteins that are linked to the translocation domain of LF [1]. To further increase cancer cell specificity, cell destructing proteins are also linked to different EGF-like domains [10]: growth factor receptors, such as the EGFR/ErbB family members, are overexpressed in malignant cells. This double-targeting approach is designed to further enhance the specificity for cancer cells.

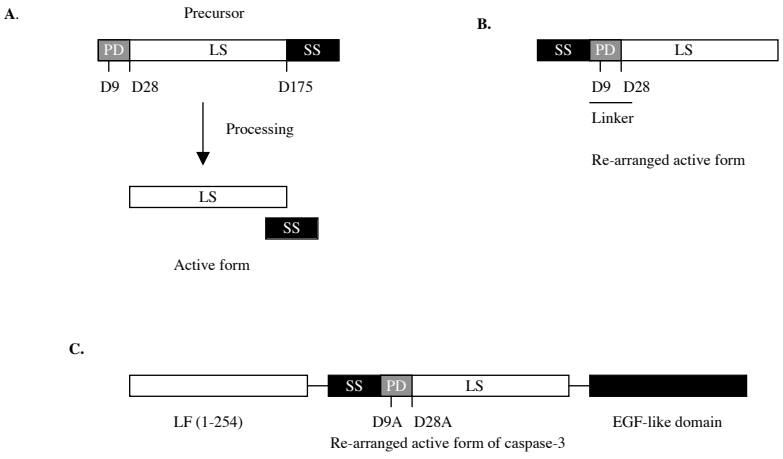

Fig. 3. Use of re-arranged active form of capases in a cellular delivery system. A. Shown is the processing of caspase-3, where the two subunits (LS and SS) in the active form of the enzyme are marked. Also marked are the aspartate (D) processing sites; PD, prodomain. B. Re-arranged, constitutively active form of caspase-3 that does not require processing. C. Hybrid protein for cellular delivery tests. The aspartate processing sites in the linker region of the re-arranged capspase-3 will be mutated. The hybrid protein also contains the domain of anthrax LF (residues 1-254) that is sufficient for cellular translocation [1] and the EGF-like domain for cancer cell targeting. Not shown are protein tags that can facilitate protein purification. Diagrams not drown to scale.

Caspases can directly cause cell destruction through protein degradation and programmed cell death [11]. However, there are several challenges to the use of these proteins in the cellular delivery system outlined above. These enzymes are expressed as inactive precursors until they are specifically activated through proteolytic processing and, moreover, the activation of several of these enzymes (e.g., caspase-2, -8, -9 and -10) requires coordinated actions of additional cellular co-factors [11, 12]. These issues have complicated our original design of using caspase-9 and hampered our progress toward successfully establishing a cellular delivery system for cell destruction. We will investigate an

alternative possibility of using activated forms of other caspases. It has been shown [13] that caspase-6 and -3 can be expressed as re-arranged enzymes that are constitutively active without requiring proteolytic processing (Fig. 3A, B). We will generate constructs to express engineered proteins (Fig. 3C) and test their delivery to cancer cells in cell-based tests. In these constructs, the aspartate processing sites in the linker regions of the re-arranged caspases will be mutated to prevent auto-cleavage; these mutations have been shown not to affect the caspase enzyme activity in degrading target proteins [13]. Once the tests of the cellular delivery system are completed and its feasibility demonstrated in cell-based assays, it can lead to future tests in whole-animal systems. It should be noted that, although our design relies on uPA/uPAR and EGFR/ErbB as cancer cell markers, other markers can also be tested in the future to efficiently and specifically eradicate breast cancer.

KEY RESEARCH ACCOMPLISHMENTS

--Further understanding of molecular principles determining biological specificity
--Analysis of protein-DNA interactions in biological specificity
--Analysis of protein-protein interactions in biological specificity/activity
--Interaction between an anti-cancer drug target (Hsp90) and a developmental protein (Bcd)
--Identifying challenges toward design of effective cancer therapeutics
--Concept of a double-targeting approach for increased specificity
--Selection of constitutively active forms (re-arranged) of caspases for cellular delivery

REPORTING OUTCOMES

Ma, J. (2005). Crossing the line between activation and repression. Trends in Genetics *21*, 54-59.

Chaney, B. A., Clark-Baldwin, K., Dave, V., Ma, J., and Rance, M. (2005). Solution structure of the K50 class homeodomain PITX2 bound to DNA and implications for mutations that cause Rieger syndrome. Biochemistry *44*, 7497-7511.

Fu, D., and Ma, J. (2005). Interplay between positive and negative activities that influence the role of Bicoid in transcription. Nucleic Acids Research *33*, 3985-3993.

Baird-Titus, J., Clark-Baldwin, K., Dave, V., Caperelli, C. A., Ma, J., and Rance, M. (2006). The Solution structure of the native K50 Bicoid homeodomain bound to the consensus TAATCC DNA-binding site. J. Molecular Biology *356*, 1137-1151.

Ma, J. (2006). Transcriptional activators and activation mechanisms. In: *Gene Expression and Regulation* (Ed. J. Ma, Higher Education Press & Springer, Beijing-New York), in press.

Ma, J. (Ed.) (2006). *Gene Expression and Regulation*, A Current Scientific Frontiers Book, Higher Education Press & Springer (Beijing-New York), in press.

Wen, Y., Fu., D., Xie, G., Wang, W., and Ma, J. (2006). Rescue of *Drosophila* Bicoid functions by a hybrid protein containing the human PITX2 homeodomain (in preparation).

CONCLUSIONS

The understanding of the molecular principles governing biological specificities represents not only a basic scientific problem but also a critical task in conquering human diseases including breast and other cancers. Our recent finding that Bcd and Hsp90 can interact with each other establishes a foundation for further investigating the molecular actions of Hsp90 and its specific inhibitor, the anti-cancer drug geldanamycin. The knowledge in biological specificities can also aid the design of therapeutics of treating cancers including the use of cellular delivery systems with a double-targeting mechanism. Our long-term objective is to understand the actions of existing anti-cancer drugs and to investigate new concepts leading to the development of new drugs to eradicate breast cancer.

REFERENCES

1. Arora, N. and S.H. Leppla, *Residues 1-254 of anthrax toxin lethal factor are sufficient to cause cellular uptake of fused polypeptides.* J Biol Chem, 1993. **268**(5): p. 3334-41.
2. Whitesell, L. and S.L. Lindquist, *HSP90 and the chaperoning of cancer.* Nat Rev Cancer, 2005. **5**(10): p. 761-72.
3. van der Straten, A., C. Rommel, B. Dickson, and E. Hafen, *The heat shock protein 83 (Hsp83) is required for Raf-mediated signalling in Drosophila.* Embo J, 1997. **16**(8): p. 1961-9.
4. Ephrussi, A. and D.S. Johnston, *Seeing is believing. The bicoid morphogen gradient matures.* Cell, 2004. **116**(2): p. 143-52.
5. Houchmandzadeh, B., E. Wieschaus, and S. Leibler, *Establishment of developmental precision and proportions in the early Drosophila embryo.* Nature, 2002. **415**(6873): p. 798-802.
6. Prince, T. and R.L. Matts, *Definition of protein kinase sequence motifs that trigger high affinity binding of Hsp90 and Cdc37.* J Biol Chem, 2004. **279**(38): p. 39975-81.
7. Petosa, C., R.J. Collier, K.R. Klimpel, S.H. Leppla, and R.C. Liddington, *Crystal structure of the anthrax toxin protective antigen.* Nature, 1997. **385**(6619): p. 833-8.
8. Liu, S., T.H. Bugge, and S.H. Leppla, *Targeting of tumor cells by cell surface urokinase plasminogen activator-dependent anthrax toxin.* J Biol Chem, 2001. **276**(21): p. 17976-84.
9. Andreasen, P.A., R. Egelund, and H.H. Petersen, *The plasminogen activation system in tumor growth, invasion, and metastasis.* Cell Mol Life Sci, 2000. **57**(1): p. 25-40.

10. Yang, D., C.T. Kuan, J. Payne, A. Kihara, A. Murray, L.M. Wang, M. Alimandi, J.H. Pierce, I. Pastan, and M.E. Lippman, *Recombinant heregulin-Pseudomonas exotoxin fusion proteins: interactions with the heregulin receptors and antitumor activity in vivo.* Clin Cancer Res, 1998. **4**(4): p. 993-1004.
11. Fadeel, B. and S. Orrenius, *Apoptosis: a basic biological phenomenon with wide-ranging implications in human disease.* J Intern Med, 2005. **258**(6): p. 479-517.
12. Acehan, D., X. Jiang, D.G. Morgan, J.E. Heuser, X. Wang, and C.W. Akey, *Three-dimensional structure of the apoptosome: implications for assembly, procaspase-9 binding, and activation.* Mol Cell, 2002. **9**(2): p. 423-32.
13. Srinivasula, S.M., M. Ahmad, M. MacFarlane, Z. Luo, Z. Huang, T. Fernandes-Alnemri, and E.S. Alnemri, *Generation of constitutively active recombinant caspases-3 and -6 by rearrangement of their subunits.* J Biol Chem, 1998. **273**(17): p. 10107-11.

Published online July 19, 2005

Nucleic Acids Research, 2005, Vol. 33, No. 13 ***3985–3993***
doi:10.1093/nar/gki691

Interplay between positive and negative activities that influence the role of Bicoid in transcription

Dechen Fu and Jun Ma*

Division of Developmental Biology, Cincinnati Children's Hospital Research Foundation, Graduate Program in Molecular and Developmental Biology, University of Cincinnati College of Medicine, 3333 Burnet Avenue Cincinnati, OH 45229, USA

Received April 1, 2005; Revised May 20, 2005; Accepted June 16, 2005

ABSTRACT

The *Drosophila* mophogenetic protein Bicoid (Bcd) can activate transcription in a concentration-dependent manner in embryos. It contains a self-inhibitory domain that can interact with the co-repressor Sin3A. In this report, we study a Bcd mutant, Bcd(A57–61), which has a strengthened self-inhibitory function and is unable to activate the *hb-CAT* reporter in *Drosophila* cells, to analyze the role of co-factors in regulating Bcd function. We show that increased concentrations of the co-activator dCBP in cells can switch this protein from its inactive state to an active state on the *hb-CAT* reporter. The C-terminal portion of Bcd(A57–61) is required to mediate such activity-rescuing function of dCBP. Although capable of binding to DNA *in vitro*, Bcd(A57–61) is unable to access the *hb* enhancer element in cells, suggesting that its DNA binding defect is only manifested in a cellular context. Increased concentrations of dCBP restore not only the ability of Bcd(A57–61) to access the *hb* enhancer element in cells but also the occupancy of the general transcription factors TBP and TFIIB at the reporter promoter. These and other results suggest that an activator can undergo switches between its active and inactive states through sensing the opposing actions of positive and negative co-factors.

INTRODUCTION

Regulation of gene transcription plays a critical role in many biological processes that range from cell growth and differentiation to embryonic patterning (1,2). Genes that participate in these biological processes need to be specifically turned on or off by transcription factors at the appropriate time and location. It is becoming increasingly clear that many transcription factors can act as both activators and repressors in a context-dependent manner [reviewed in (3)]. Promoter/enhancer architecture and cellular levels of other proteins have been suggested to play roles in influencing a transcription factor's regulatory functions, but the precise mechanisms in most cases remain largely unclear. For proteins that can work as both activators and repressors, they have three distinct activity states: active, repressive and inactive (neither active nor repressive). In contrast, for proteins that work only as activators, such as the *Drosophila* protein Bicoid (Bcd), they only have two activity states: active and inactive. Analysis of these proteins can thus help us understand the important question of how the simple on–off switches of activator activities are achieved. Bcd is a well-documented protein that undergoes such on–off activity switches in a concentration-dependent manner (see below). The experiments described here suggest another mechanism in which the opposing actions of positive and negative co-factors can facilitate Bcd to switch between its active and inactive states in a manner that is independent of Bcd concentration.

Bcd is a molecular morphogen that plays a critical role in patterning embryonic structures, including the head and thorax (4,5). This 489 amino acids transcription factor contains a homeodomain (residues 92–151) in its N-terminal portion (6). Bcd, which is distributed in the early embryo as an anterior-to-posterior gradient, is responsible for activating specific target genes in a concentration-dependent manner. For example, *orthodenticle* (*otd*), *hunchback* (*hb*) and *knirps* (*kni*) are direct Bcd target genes that are required for patterning the head, thoracic and abdominal structures, respectively (7). These genes are expressed in distinct parts of the embryo by responding to different Bcd concentrations (8–10). Bcd has the ability to bind DNA in a highly cooperative manner (11–14), and it has been suggested that the affinity of Bcd binding sites in an enhancer can determine the concentration of Bcd required

*To whom correspondence should be addressed. Tel: +1 513 636 7977; Fax: +1 513 636 4317; Email: jun.ma@cchmc.org
Present address:
Dechen Fu, Department of Molecular and Cellular Biology, University of California, Berkeley, CA 94720, USA

for activating transcription (9,15,16). Our recent studies suggest that the arrangements of Bcd binding sites in an enhancer can also play a critical role in regulating the activity of Bcd and contributing to its concentration-dependent action (17).

CBP is a co-activator that interacts with many transcription factors and participates in the activation process (18,19). Its histone acetyltransferase (HAT) enzymatic activity is thought to alter chromatin structure by acetylating the histone tails thus increasing the accessibility of DNA for both gene-specific transcription factors and general transcription factors (GTFs). CBP can also play a structural role by bridging between transcription factors and GTFs or by recruiting other HAT activities (18,19). In *Drosophila*, dCBP has been shown to be a co-activator for Ci (20), Mad (21) and Dorsal (22). dCBP also plays a role in facilitating Bcd to activate transcription (23). dCBP and Bcd can interact with each other through distinct domains on different enhancers. In particular, on the *hb* enhancer element the C-terminal portion of Bcd plays an important role in responding to the co-activation function of dCBP, whereas on the *kni* enhancer element, the N-terminal domain plays an important role (23).

In addition to its ability to interact with co-activators, such as dCBP, Bcd can also interact with co-repressors. An analysis of the N-terminal region of Bcd revealed a self-inhibitory domain (residues 52–91) that can dramatically inhibit the ability of Bcd to activate transcription (24). For example, on the *hb-CAT* reporter gene which contains the Bcd-responsive *hb* enhancer element, a Bcd derivative lacking the entire N-terminal domain, Bcd(92–489), exhibits an activity 40 times higher than the full-length protein in *Drosophila* S2 cells. A systematic analysis of the self-inhibitory domain identified a 10 amino acid motif (residues 52–61) that is most critical for the self-inhibitory function. Interestingly, mutations of different residues in this motif can cause drastically opposing effects (25). In particular, the mutant protein Bcd(A52–56), which has residues 52–56 changed to alanines, is 25 times more active than wt Bcd on the *hb-CAT* reporter in S2 cells. In contrast, on the same reporter another mutant, Bcd(A57–61), which has the neighboring five amino acids changed to alanines, is virtually inactive (<2% of wt Bcd activity) at all concentrations. The co-repressor Sin3A has been shown to interact with the evolutionarily conserved N-terminal domain of Bcd, and it is proposed that mutations that alter the 10 amino acid motif can weaken or strengthen this interaction, thus increasing or decreasing, respectively, the activity of Bcd (25). Another component of the Sin3A-HDAC (histone deacetylase) complex, SAP18, has also been shown to interact with Bcd, apparently through multiple Bcd domains [(24,26); see Figure 1A for a schematic diagram of Bcd domains interacting with co-factors].

In this report, we use Bcd(A57–61), an inactive protein on the *hb-CAT* reporter in S2 cells, as a tool to analyze

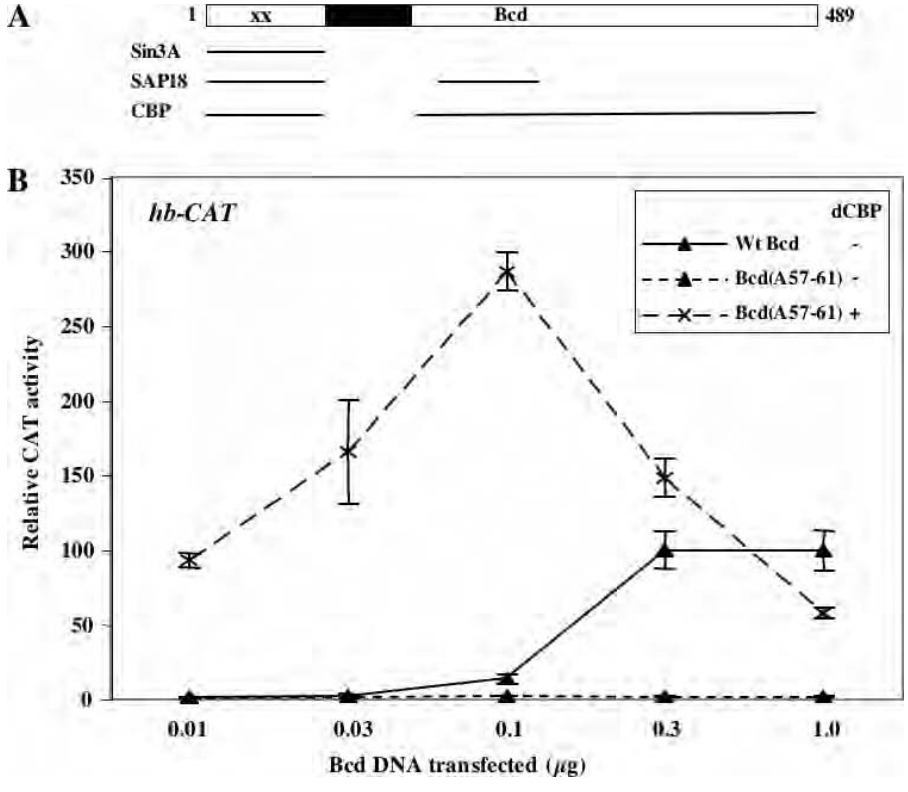

Figure 1. Exogenous dCBP switches the activity states of Bcd(A57–61) in S2 cells. (**A**) Shown is a schematic diagram of Bcd and its interacting domains with co-factors. The homeodomain (residues 92–151) of the 489 amino acid Bcd protein is marked with a black box and the two neighboring mutations discussed in this report, A52–56 and A57–61, are each marked with an 'X'. The interaction information in this diagram is based on (25) for Sin3A, (24,26) for SAP18 and (23) for CBP. The diagram is not drawn to scale. (**B**) Shown are CAT assay results in S2 cells that were transfected with the reporter plasmid *hb-CAT* (1 μg), the indicated amounts of effector plasmids expressing wt Bcd or Bcd(A57–61), with (+) or without (−) another effector plasmid (5 μg) expressing dCBP. Fold activation by wt Bcd (at 1 μg transfected DNA) without exogenous dCBP was set to 100.

the interplay between positive and negative co-factors in regulating Bcd function. This mutant exhibits some special properties. In particular, while it is inactive on the *hb-CAT* reporter gene, it can activate another reporter gene, *kni-CAT* (17). These and other findings suggest that the activity state of this protein is intricately controlled, and we sought to gain a better understanding of this mutant protein by focusing on the roles of, and the interplay between, Bcd interacting co-factors. In this report, we show that increased concentrations of dCBP in S2 cells can switch this protein from an inactive state to an active one on the *hb-CAT* reporter. We further show that the C-terminal domain of Bcd(A57–61) mediates such activity-rescuing function of dCBP. We provide evidence demonstrating that, despite its normal DNA binding ability *in vitro*, Bcd(A57–61) fails to occupy the *hb* enhancer element in cells. High levels of dCBP in S2 cells restore the ability of Bcd(A57–61) to access the *hb* enhancer element and enable this Bcd derivative to recruit the GTFs TBP and TFIIB to the target promoter. We also provide evidence suggesting that dCBP may negatively affect the interaction between Bcd and Sin3A in cells. Together, these results demonstrate that dCBP plays an important role in regulating Bcd function in a dCBP concentration-dependent manner. They suggest that the opposing actions of positive and negative co-factors can facilitate Bcd to switch between its active and inactive states in a manner that is independent on Bcd concentration.

MATERIALS AND METHODS

Plasmid construction

Plasmids expressing Bcd derivatives were generated in two steps as described previously (23). The *bcd* gene was first modified on pFY441, a pGEM3-based plasmid containing wt *bcd* linked to the coding sequence of the hemagglutinin (HA) tag, and then transferred to pFY442, a plasmid expressing HA-tagged wt Bcd from the *Drosophila actin 5C* promoter (24). For Bcd(1–246; A57–61), the pGEM3-based plasmid was pDF333 and the expression plasmid was pFD347. Reporter genes and the effector plasmids pFY443 [Bcd(1–246)] and pFY465 [Bcd(A57–61)] have been described previously (14,24). The expression plasmids of wt and mutant dCBP were kindly provided by Dr S. Smolik.

Transient transfection assays

Drosophila S2 cells were transfected with plasmids by the calcium phosphate co-precipitation method as described by Invitrogen. The total amount of DNA in each transfection was adjusted to 10 μg by salmon sperm DNA. In order to monitor the transfection efficiency, 1 μg control plasmid *pCopia-lacZ* was co-transfected in each experiment, and both CAT assays and western blot analyses were normalized according to the β-galactosidase activity. CAT activity was measured as previously described by using three independently transfected samples for each experiment (14). The protein levels of Bcd derivatives with or without dCBP were detected by western blot using anti-HA antibody (1:500 final dilution, Babco). Double-strand RNA against endogenous dCBP in S2 cells was generated as described previously (23), and the RNAi treatment did not affect the accumulation of Bcd in S2 cells.

Gel shift assays

The *hb* enhancer probe for gel shift experiments was released from a plasmid and filled-in with Klenow in the presence of [α-^{32}P]dCTP as described previously (17). Wild-type Bcd and its derivative used in these assays were expressed *in vitro* by using the TnT quick-coupled transcription/translation system (Promega). The experimental procedures and conditions for gel shift assays were described previously (17).

Chromatin-immunoprecipitation (ChIP) assays

ChIP assays were performed according to Fu *et al*. (23). The presence of Bcd, GTFs and acetylated histones at the *hb* enhancer-core promoter region was detected by PCR using primers *hb*-core5 and *hb-CAT*3 as described previously (23).

Co-immunoprecipitation (Co-IP)

Co-IP experiments were performed as described previously (23). Briefly, nuclear extracts prepared from S2 cells were incubated with anti-HA antibody (1:100 final dilution) in IP buffer (20 mM Tris–HCl, pH 8.0, 160 mM MgCl2, 0.1% Nonidet P-40 and 10% glycerol). The precipitated products were resolved by SAS–PAGE gel and detected by western blot using anti-Sin3A antibodies [kindly provided by Drs Lori Pile and David Wassarman (27)]. Quantitation of the co-IP data shown in Figure 5 was conducted as follows. The intensities of input and co-IP Sin3A bands for each sample were measured to obtain an intensity ratio of co-IP product over input. For each experiment, the ratio for Bcd transfection alone (lane 4) was arbitrarily set to 100 to allow comparison of data from independent experiments.

RESULTS

High levels of dCBP switch Bcd(A52–61) to an active state

Our previous transfection experiments in S2 cells have shown that Bcd(A57–61) has a strengthened self-inhibitory function and is nearly completely inactive on the *hb-CAT* reporter gene (17,25). This mutant protein is stably accumulated in cells (25), suggesting that its inability to activate *hb-CAT* reflects a distinct functional state of this protein rather than its defects in protein stability. Unlike wt Bcd, which exhibited a dose-dependent activation function in transfection assays (Figure 1B, solid line), this mutant protein failed to activate *hb-CAT* at all concentrations tested (Figure 1B, dashed line, bottom). To determine whether high concentrations of the co-activator dCBP might counteract the strengthened self-inhibitory function and switch Bcd(A57–61) to an active state, we conducted co-transfection experiments. In these experiments, the ability of Bcd(A57–61) to activate *hb-CAT* was measured in the presence or absence of dCBP exogenously expressed from a transfected plasmid.

Our results showed that exogenous dCBP dramatically rescued the activity of Bcd(A57–61) on *hb-CAT*, increasing its activity by 29 to 110 fold depending on Bcd concentration (Figure 1B, dashed line, top; also see Table 1). In the presence of exogenous dCBP, the activity of Bcd(A57–61) at several concentrations was higher than wt Bcd at its saturating concentrations (without exogenous dCBP). Table 1 lists the effect

of dCBP on wt Bcd and Bcd(A57–61), further indicating that Bcd(A57–61) responds to dCBP much more robustly than wt Bcd does at all concentrations tested. As shown previously, exogenous dCBP has no effect on reporter gene expression in the absence of Bcd and does not alter the amount of Bcd protein in cells (23). Together, these results suggest that dCBP is a limiting co-factor for Bcd(A57–61) and is capable of making this Bcd protein to switch between its inactive and active states on the *hb-CAT* reporter in cells.

Rescue of Bcd(A56–61) activity by dCBP requires the C-terminal domain of Bcd

It has been shown that Bcd and dCBP can physically interact with each other (23). Deletion analysis further suggested that the C-terminal half of Bcd plays an important role in responding to the co-activator function of dCBP on the *hb-CAT* reporter (23). To determine whether this domain is required for mediating the activity-rescuing function of dCBP, we analyzed the effect of exogenous dCBP on a truncated derivative of Bcd, Bcd(1–246). Two versions of Bcd(1–246), with either wt or the A57–61 mutation at its N-terminus, were used in the experiments. As shown previously (23), the truncated derivative Bcd(1–246) responded to dCBP modestly (Figure 2A). However, dCBP failed to rescue the activity of the truncated, mutant protein Bcd(1–246; A57–61) at all concentrations tested (Figure 2A). dCBP did not affect the accumulated levels of the Bcd proteins (Figure 2B). These results suggest that dCBP rescues the activity of Bcd(A57–61) through the C-terminal domain of Bcd.

Table 1. The effect of dCBP on wt Bcd and Bcd(A57–61)

DNA transfected (μg)	Effect of dCBP (fold increase)	
	wt Bcd	Bcd(A57–61)
0.01	17	58
0.03	19	72
0.1	4.8	110
0.3	4.7	61
1.0	7.5	29

Listed is the effect (fold increase) of dCBP on wt Bcd and Bcd(A57–61) in activating the *hb-CAT* reporter gene in S2 cells. The amount of transfected DNA refers to the plasmids expressing the Bcd derivatives. The data for wt Bcd and Bcd(A57–61) are from Fu *et al*. (23) and Figure 1B, respectively.

Defect of Bcd(A57–61) in *hb* enhancer recognition in cells but not *in vitro*

Bcd(A57–61) has a normal ability to bind to a single TAATCC site when analyzed *in vitro* (25). To determine whether this mutant protein might be defective in recognizing natural enhancer elements that contain multiple Bcd binding sites, we conducted gel shift studies using the *hb* enhancer element. As shown previously (17), wt Bcd bound to this enhancer element in a cooperative manner, forming protein–DNA complexes that contained multiple Bcd molecules (Figure 3, lanes 1–4). Our gel shift experiments using Bcd(A57–61) showed that this mutant protein can bind to the *hb* enhancer element in a manner comparable with the wt Bcd protein

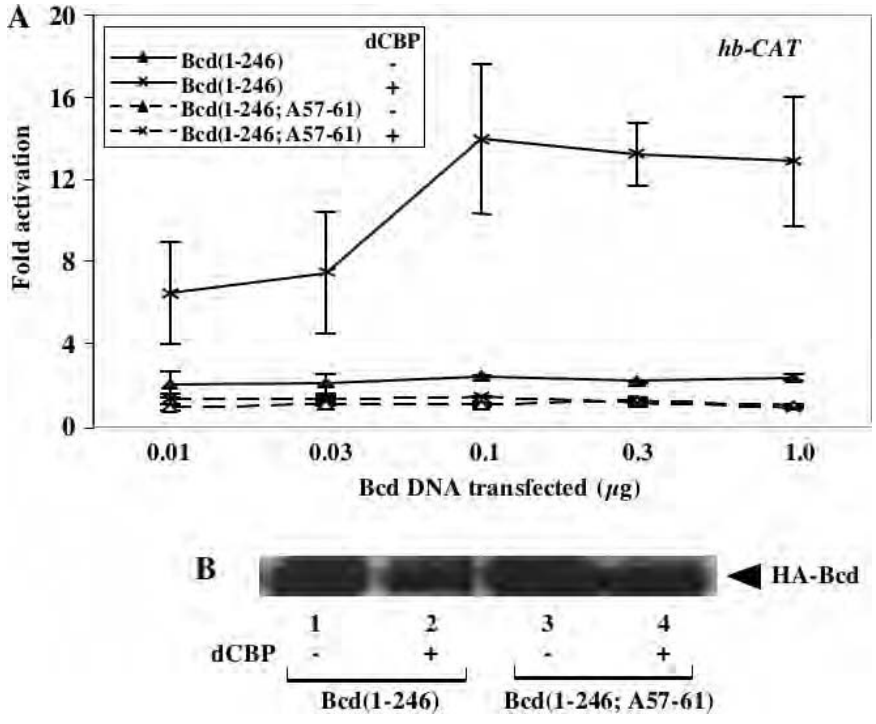

Figure 2. Switch of Bcd(A57–61) activity states by dCBP requires Bcd C-terminal domain. (**A**) Shown are CAT assay results in S2 cells that were transfected with the reporter plasmid *hb-CAT* (1 μg), the indicated amounts of effector plasmids expressing two different Bcd derivatives, with (+) or without (−) the effector plasmid (5 μg) expressing dCBP. The two Bcd derivatives are Bcd(1–246), a truncated Bcd with a wt N-terminus; Bcd(1–246; A57–61), a truncated derivative with the A57–61 mutation in its self-inhibitory domain. Fold activation for each assay, measured by CAT activity, is shown in the figure. (**B**) Western blot data showing the HA-tagged Bcd protein levels (1 μg transfected DNA) in the presence (+) or absence (−) of dCBP (5 μg transfected DNA).

(Figure 3, lanes 5–8). Bcd(A57–61) also bound to another natural enhancer element, *kni*, in a cooperative manner similar to the wt Bcd protein *in vitro* (data not shown). These results further support our conclusion that the A57–61 mutation of Bcd does not abolish the protein's ability to recognize DNA *in vitro* (25).

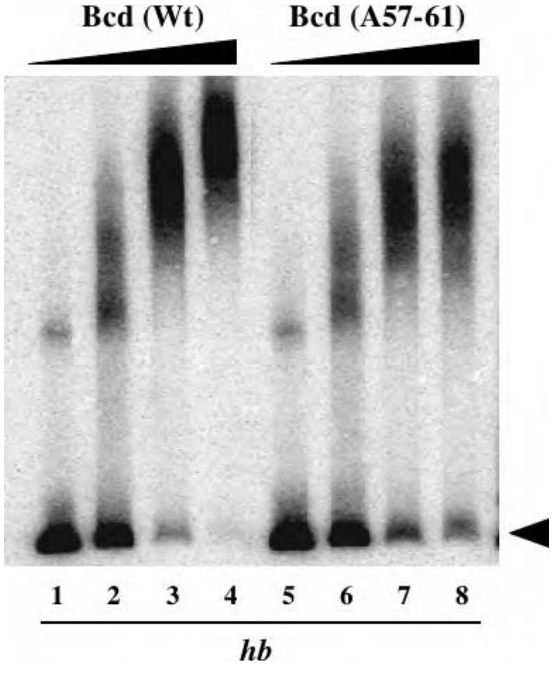

Figure 3. Enhancer element binding by Bcd(A57–61) *in vitro*. Gel shift data showing DNA binding to the *hb* enhancer element by wt Bcd (lanes 1–4) and Bcd(A57–61) (lanes 5–7). Free probe is indicated by an arrowhead (bottom right). In the absence of Bcd, there were no shifted complexes detected (data not shown).

To further dissect the defects of Bcd(A57–61), thus helping understand the mechanisms of the functional rescue by dCBP, we carried out a ChIP analysis in cells. We compared the occupancy of wt Bcd and Bcd(A57–61) at the *hb-CAT* reporter. We specifically chose conditions in which Bcd proteins were expressed at high levels to reveal functional defects of the mutant Bcd that could not be overcome by increased Bcd concentrations (also see Figure 1B for reporter assay data). As shown previously (23), our ChIP experiments detected a significant occupancy of wt Bcd at the *hb* enhancer element of the reporter gene [Figure 4B (a), lane 9]. In contrast, Bcd(A57–61) failed to exhibit an occupancy above background levels at the *hb* enhancer element in the same ChIP assays [Figure 4B (a), lane 11]. Together, these results suggest that Bcd(A57–61), despite its normal ability to bind DNA *in vitro*, has a functional defect in accessing the *hb* enhancer element in cells.

dCBP restores the occupancy of Bcd(A57–61) at *hb* enhancer in cells

To determine whether dCBP can affect the ability of Bcd(A57–61) to access the *hb-CAT* reporter gene in cells, we conducted ChIP experiments in the presence of exogenously expressed dCBP (see Figure 4A for a schematic diagram of the reporter gene). As shown by the ChIP data [Figure 4B (a), lanes 11 and 12], dCBP restored the occupancy of Bcd(A57–61) at the *hb* enhancer element in cells [Figure 4B (a), lane 12]. Under the conditions of high Bcd concentrations, dCBP had little effect on wt Bcd (lanes 9 and 10) as shown previously (23).

Our ChIP experiments also revealed a restored occupancy of GTFs at the *hb-CAT* reporter caused by high levels of dCBP. In the absence of exogenous dCBP, Bcd(A57–61) failed to enhance the occupancy of either TBP or TFIIB at the promoter region [Figure 4B (b and c), compare lanes 7, 9 and 11]. In the presence of exogenous dCBP, Bcd(A57–61) increased the

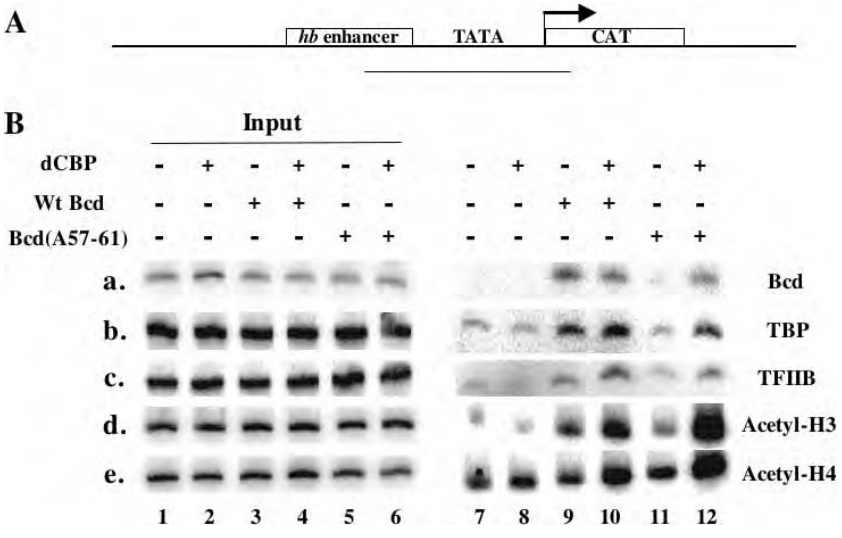

Figure 4. Restored occupancy of Bcd(A57–61) and GTFs by exogenous dCBP. (**A**) Shown is a schematic diagram of the reporter gene, marking the promoter region (thin line) used for detection by PCR in ChIP assays. The diagram is not drawn to scale. (**B**) Shown are ChIP data from S2 cells that were transfected with plasmids expressing the indicated effectors [dCBP proteins, 5 μg; Bcd(A57–61), 1 μg] and the *hb-CAT* reporter plasmid (1 μg). Antibodies used for ChIP assays were: HA to detect HA-tagged Bcd (panel a), TBP (panel b), TFIIB (panel c), acetyl-H3 (panel d) and H4 (panel e). Lanes 1–6 show input controls, which represent the PCR product of 1% of the total isolated DNA used in the ChIP assays.

occupancy of both TFIIB and TBP (lane 12). Finally, our ChIP experiments showed that dCBP increased the acetyl-H3 and H4 levels at the reporter in the presence of either wt Bcd or Bcd(A57–61) [Figure 4B (d and e), lanes 9–12]. In all the cases, the effects of dCBP required the presence of Bcd or its derivative (compare lanes 7 and 8), indicating that these observed effects represent Bcd-dependent functions of dCBP. Together, these results reveal not only a restored occupancy, caused by increased dCBP levels in cells, of Bcd(A57–61) at the *hb-CAT* reporter but also an elevated recruitment of GTFs and an increased histone acetylation level at the reporter.

dCBP may negatively affect Bcd–Sin3A interaction in cells

As further detailed in Discussion (below), several models are consistent with our finding that high levels of dCBP can restore activity to Bcd(A57–61). For example, it is possible that dCBP and Sin3A may compete for Bcd interaction, thus representing antagonistic forces to influence Bcd function. To determine whether the interaction between Bcd and Sin3A might be affected by dCBP, we conducted co-IP experiments in cells with altered dCBP levels. To reduce cellular levels of dCBP, we used an RNAi approach, which has been shown to specifically affect Bcd activity without altering the amount of Bcd in cells (23). We used exogenously expressed dCBP to increase its cellular levels. As shown in our co-IP experiments (Figure 5A), the amount of Sin3A precipitated by Bcd was increased by dCBP RNAi treatment (lane 6) and marginally affected by dCBP overexpression (lane 5). Figure 5B shows the quantitation of the data from three independent experiments (relative amounts of co-IP Sin3A for lanes 4, 5 and 6 are 100, 87 ± 16 and 276 ± 88, respectively). These results suggest that dCBP may negatively affect the interaction between Bcd and Sin3A. The modest effect of dCBP on Sin3A–Bcd interaction suggests that such an antagonistic effect may represent only one of the several individually weak mechanisms by which dCBP rescues the activity of Bcd(A57–61) (see Discussion for further details).

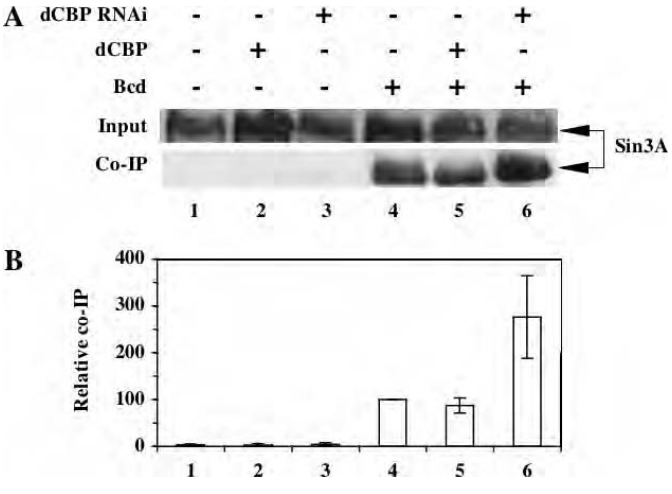

Figure 5. Interaction between Bcd and Sin3A may be affected by dCBP. (**A**) The interaction between HA-tagged Bcd and the endogenous Sin3A in S2 cells was detected by a co-IP analysis (see Materials and Methods for details). Sin3A co-precipitated by anti-HA antibodies was detected by western blot using anti-Sin3A antibodies. S2 cells were either transfected (+) or not (−) with the indicated plasmids expressing HA-Bcd (1 μg) and dCBP (5 μg), and had either been subject (+) or not (−) to dCBP RNAi treatment (25 μg dsRNA). In this figure, lanes 1–3 are controls showing that no Sin3A was precipitated in the absence of HA-Bcd. Input represents one-tenth of total nuclear extract used in the co-IP assay as described previously (23). (**B**) The relative amounts of the co-IP Sin3A product in three independent experiments were quantified (see Materials and Methods) and the results are shown (mean ± SD); all lanes in this graph correspond to those in (A).

The HAT-deficient mutant dCBP can partially rescue Bcd(A57–61) activity

Our previous experiments have shown that dCBP can increase the activity of wt Bcd through both HAT-dependent and -independent mechanisms (23). A HAT-independent action of dCBP suggests a structural role of this protein in regulating Bcd activity, a suggestion consistent with the observed negative effect of dCBP on Bcd-Sin3A interaction (Figure 5). To specifically determine whether the HAT activity of dCBP is required for its ability to restore function to Bcd(A57–61) on the *hb-CAT* reporter, we used a HAT-deficient mutant of dCBP (28); both wt dCBP and this mutant protein are accumulated to similar levels when expressed in S2 cells (23). As shown in Figure 6, this mutant dCBP increased partially the activity of wt Bcd on the *hb-CAT* reporter [lanes 4–6; also see (23)]. It also rescued, though with a reduced efficiency, the activity of Bcd(A57–61) on the *hb-CAT* reporter (lanes 7–9). Together, these results suggest that dCBP can play an enzyme activity-independent role in rescuing the activity of Bcd(A57–61) on the *hb-CAT* reporter in cells.

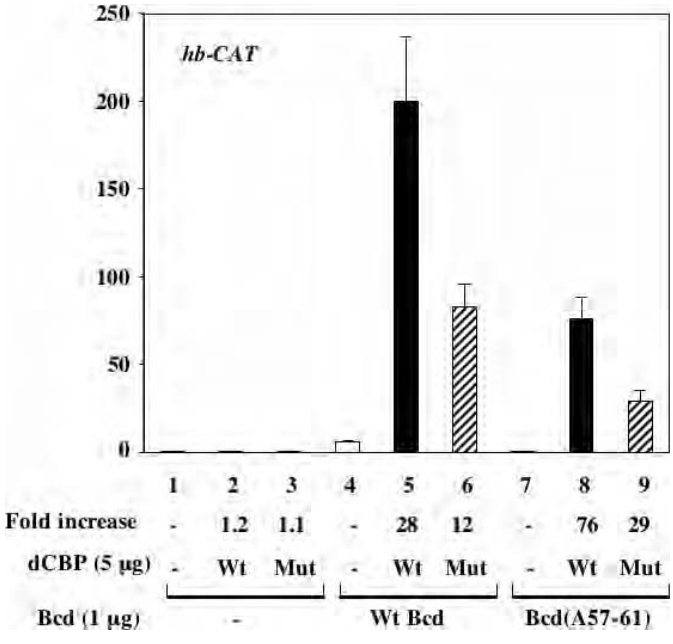

Figure 6. HAT-deficient dCBP can partially rescue the activity of Bcd(A57–61). Shown are CAT assay results in S2 cells that were transfected with the reporter plasmid *hb-CAT* (1 μg), the effector plasmids (1 μg) expressing the indicated Bcd proteins, with (+) or without (−) another effector plasmid (5 μg) expressing dCBP. The increase of Bcd activity by wt and mutant dCBP proteins is also indicated in the figure as fold increase. The exogenously expressed wt and mutant dCBP proteins are accumulated at similar levels in S2 cells as described previously (23).

DISCUSSION

As a molecular morphogen, Bcd can undergo switches, in a concentration-dependent manner, between its active and inactive states in activating transcription of its target genes. The experiments described in this report suggest another mechanism that can facilitate on–off switches of Bcd activity in a Bcd concentration-independent manner. In particular, the mutant Bcd(A57–61) is incapable of activating the *hb-CAT* reporter gene in S2 cells at all concentrations tested (Figure 1B). The inability of this mutant Bcd to activate the *hb-CAT* reporter reflects a distinct functional state of this protein rather than its defects in protein stability. In fact, this same mutant protein is only modestly weaker than the wt protein on another reporter gene, *kni-CAT*, which contains the Bcd-responsive *kni* enhancer element (17). These and other results suggested that the A57–61 mutation may cause its functionally inactive state on *hb-CAT* by more efficiently interacting with a co-repressor protein(s), such as Sin3A and its associated complex(es) (24,25). The experiments described in this report show that increased concentrations of dCBP can restore activity to Bcd(A57–61) on the *hb-CAT* reporter in cells. These results suggest that the opposing actions of positive and negative co-factors can facilitate Bcd to switch between its active and inactive states in a manner that is Bcd concentration-independent.

Although Bcd(A57–61) can bind to both a single site and natural enhancer elements *in vitro*, it is unable to access the *hb* enhancer element in cells (Figures 3 and 4). These results suggest that the DNA binding defect of this mutant protein is only manifested in a cellular context. This notion is consistent with our finding that the PAH domains of Sin3A do not exhibit any increased ability to reduce DNA binding by Bcd(A57–61) *in vitro* when compared with wt Bcd (data not shown) (25). We propose that other co-repressors or those that are associated with Sin3A, such as the HDACs, can reduce the ability of Bcd to access a natural enhancer in cells. It is possible that the enzymatic HDAC activity that is more stably associated with Bcd(A57–61) makes it unable to negotiate with histones for accessing DNA. It is also possible that a more stable Bcd-co-repressor complex may sterically hinder the interaction between Bcd(A57–61) molecules and prevent cooperative binding to the enhancer element in cells.

The most striking finding of this report is that high levels of dCBP can switch Bcd(A57–61) from its inactive state to an active one on the *hb-CAT* reporter in cells. Our ChIP data further show that dCBP increases both the ability of Bcd(A57–61) to access the *hb* enhancer element in cells and the occupancy of GTFs at the reporter promoter (Figure 4B). How does dCBP switch the activity states of Bcd(A57–61) on *hb-CAT* in cells? Since Bcd and dCBP can physically interact with each other through multiple domains (23) (Figure 1A), it is possible that dCBP may increase the DNA binding ability of Bcd in cells by stabilizing the interaction between Bcd molecules and thus enhancing its cooperativity. It is also possible that dCBP may physically compete with co-repressor complexes in interacting with Bcd. Our co-IP results suggest that dCBP may negatively affect the interaction between Bcd and Sin3A in cells (Figure 5). dCBP could also play a role in facilitating the interaction between Bcd and the transcription machinery. For all these actions, dCBP may play a structural (rather than enzymatic) role (Figure 6). Finally, the fact that the HAT-defective mutant of dCBP does have a reduced ability to restore activity to Bcd(A57–61) (Figure 6) indicates that its enzymatic activity has a positive role, possibly through modifications of histones. It is likely that dCBP can affect the Bcd(A57–61) activity through multiple mechanisms that may be weak individually (Figures 5 and 6) but, when combined, can lead to a dramatic switch from its inactive state to an active one on the *hb-CAT* reporter in cells.

Currently, it is poorly understood how precisely Bcd activates transcription. Previous studies suggest that much of its activation function is conferred by the C-terminal portion of Bcd (16,29). This portion of the protein contains several domains, including the acidic, glutamine-rich and alanine-rich domains, that are characteristic of activation domains capable of interacting with components of the transcription machinery (16,29–31). Interestingly, the alanine-rich domain previously thought to play an activation role was shown recently to exhibit an inhibitory function instead (32). The C-terminal domain of Bcd can also interact with dCBP (23), and our results show that this domain is responsible for mediating the activity-switching function of dCBP (Figure 2). Although much of the activation function of Bcd is provided by its C-terminal domain, the N-terminal portion of the protein also contains some activation function. Studies have shown that Bcd(1–246), a derivative lacking the entire C-terminal portion of Bcd, can rescue the bcd^{-} phenotype when expressed at high levels (33). These results suggest that Bcd can achieve its activation function through multiple domains presumably by interacting with different proteins, including co-activators and components of the transcription machinery. The results described in this report further support the importance of dCBP in facilitating activation by Bcd.

Bcd is a morphogenetic protein whose behavior can be regulated not only by its own concentration but also by the enhancer architecture (17). Our recent experiments show that, on the *kni* and *hb* enhancer elements, the N-terminal domain of Bcd is preferentially used for either cooperative DNA binding or self-inhibition, respectively (17). We propose that the interaction between Bcd molecules bound to the *kni* enhancer element, through its N-terminal domain, can interfere with its interaction with co-repressors, such as Sin3A. As described in this report, co-activators such as dCBP and co-repressors such as Sin3A can also functionally antagonize each other, possibly by competing for Bcd interaction as part of the mechanisms (Figure 5). Bcd is more sensitive to the self-inhibitory function on the *hb* enhancer element than on the *kni* enhancer element (17); consistent with dCBP's antagonistic role, dCBP increases the activity of Bcd more robustly on the *hb* enhancer element than on the *kni* enhancer element (23). However, the interplay between positive and negative activities that regulate Bcd functions is probably far more complex than the simple physical competition: as already discussed above, dCBP can affect Bcd activity through multiple mechanisms in both HAT-dependent and independent manners (Figure 6) (23). Moreover, in the presence of exogenous dCBP, high levels of Bcd(A57–61) cause a reduction in its activity on the *hb-CAT* reporter in cells (Figure 1B), a reduction that is not observed with wt Bcd (23), suggesting that the optimal concentration ratio between Bcd and dCBP may vary depending on the strengths of the self-inhibitory function and

interaction with co-repressors. In addition, high concentrations of dCBP can rescue the inactive derivative Bcd(A57–61), but not another inactive derivative lacking the C-terminal portion, Bcd(1–246; A57–61), suggesting that the Bcd–dCBP interaction strength can also influence the balance between positive and negative activities that regulate Bcd function.

The experiments described in this report suggest that an activator's function is subject to intricate controls by both positive and negative activities in cells. A fine balance between these activities is critical for normal cellular and developmental processes. Our transgenic experiments show that both Bcd(A57–61), which has a strengthened self-inhibitory function, and Bcd(A52–56), which has a weakened self-inhibitory function, cause embryonic defects [(24) and unpublished data]. In addition, embryos with reduced dCBP activity exhibit defects in early expression patterns of a Bcd target gene, even-skipped [(23) and Y. Wen, A. York and J. Ma, unpublished data]. Finally, a recent study reveals that mutations affecting SAP18, a component of the Sin3A-HDAC complex, can alter Bcd function and anterior patterning in embryos (34). In addition to the co-factors discussed here (Sin3A, dCBP and SAP18), Bcd likely has the ability to interact with many other proteins, including not only regulatory proteins but also components of the transcription machinery (30,31). Precisely how all these different proteins harmoniously regulate and facilitate the execution of Bcd functions during development remains to be determined. Recent studies have shown that the Bcd gradient in embryos possesses a strikingly sophisticated ability to activate its target genes in a precise manner (35–37). These findings further underscore the need of intricate control mechanisms that facilitate Bcd to switch between its active and inactive states in target gene activation. Our studies suggest that on–off switches of Bcd activity can be achieved not only in a Bcd concentration-dependent manner but also in a Bcd concentration-independent manner. It remains to be investigated whether and how Bcd interacting proteins, including those yet to be identified, participate in the precision control of target gene activation during development.

ACKNOWLEDGEMENTS

The authors thank Dr S. Smolik for providing the dCBP clones used in this study, Drs Lori Pile, David Wassarman and James Kadonaga for providing necessary antibodies, and Dr D. Wiginton, members of this laboratory, and anonymous reviewers for comments on the manuscript. This work was supported in part by grants from NSF, NIH, AHA and DOD. Funding to pay the Open Access publication charges for this article was provided by NSF.

Conflict of interest statement. None declared.

REFERENCES

1. Levine,M. and Tjian,R. (2003) Transcription regulation and animal diversity. *Nature*, **424**, 147–151.
2. Kadonaga,J.T. (2004) Regulation of RNA polymerase II transcription by sequence-specific DNA binding factors. *Cell*, **116**, 247–257.
3. Ma,J. (2005) Crossing the line between activation and repression. *Trends Genet*., **21**, 54–59.
4. Driever,W. (1992) The Bicoid morphogen: concentration dependent transcriptional activation of zygotic target genes during early *Drosophila* development. In McKnight,S.L. and Yamamoto,K. (eds), *Transcriptional Regulation*. Cold Spring Harbor Laboratory Press, NY, pp. 1221–1250.
5. Ephrussi,A. and Johnston,D.S. (2004) Seeing is believing: the bicoid morphogen gradient matures. *Cell*, **116**, 143–152.
6. Berleth,T., Burri,M., Thoma,G., Bopp,D., Richstein,S., Frigerio,G., Noll,M. and Nüsslein-Volhard,C. (1988) The role of localization of bicoid RNA in organizing the anterior pattern of the *Drosophila* embryo. *EMBO J*., **7**, 1749–1756.
7. Rivera-Pomar,R. and Jackle,H. (1996) From gradients to stripes in *Drosophila* embryogenesis: filling in the gaps. *Trends Genet*., **12**, 478–483.
8. Driever,W. and Nüsslein-Volhard,C. (1988) The bicoid protein determines position in the Drosophila embryo in a concentration-dependent manner. *Cell*, **54**, 95–104.
9. Gao,Q. and Finkelstein,R. (1998) Targeting gene expression to the head: the Drosophila orthodenticle gene is a direct target of the Bicoid morphogen. *Development*, **125**, 4185–4193.
10. Rivera-Pomar,R., Lu,X., Perrimon,N., Taubert,H. and Jackle,H. (1995) Activation of posterior gap gene expression in the *Drosophila* blastoderm. *Nature*, **376**, 253–256.
11. Ma,X., Yuan,D., Diepold,K., Scarborough,T. and Ma,J. (1996) The *Drosophila* morphogenetic protein Bicoid binds DNA cooperatively. *Development*, **122**, 1195–1206.
12. Yuan,D., Ma,X. and Ma,J. (1996) Sequences outside the homeodomain of bicoid are required for protein–protein interaction. *J. Biol. Chem*., **271**, 21660–21665.
13. Burz,D.S., Pivera-Pomar,R., Jackle,H. and Hanes,S.D. (1998) Cooperative DNA-binding by Bicoid provides a mechanism for threshold-dependent gene activation in the *Drosophila* embryo. *EMBO J*., **17**, 5998–6009.
14. Zhao,C., Dave,V., Yang,F., Scarborough,T. and Ma,J. (2000) Target selectivity of bicoid is dependent on non-consensus site recognition and protein–protein interaction. *Mol. Cell. Biol*., **20**, 8112–8123.
15. Driever,W., Thoma,G. and Nüsslein-Volhard,C. (1989) Determination of spatial domains of zygotic gene expression in the *Drosophila* embryo by the affinity of binding site for the bicoid morphogen. *Nature*, **340**, 363–367.
16. Struhl,G., Struhl,K. and Macdonald,P.M. (1989) The gradient morphogen bicoid is a concentration-dependent transcriptional activator. *Cell*, **57**, 1259–1273.
17. Fu,D., Zhao,C. and Ma,J. (2003) Enhancer sequences influence the role of the amino-terminal domain of Bicoid in transcription. *Mol. Cell. Biol*., **23**, 4439–4448.
18. Goodman,R.H. and Smolik,S. (2000) CBP/p300 in cell growth, transformation, and development. *Genes Dev*., **14**, 1553–1577.
19. Chan,H.M. and La Thangue,N.B. (2001) p300/CBP proteins: HATs for transcriptional bridges and scaffolds. *J. Cell Sci*., **114**, 2363–2373.
20. Akimaru,H., Chen,Y., Dai,P., Hou,D.X., Nonaka,M., Smolik,S.M., Armstrong,S., Goodman,R.H. and Ishii,S. (1997) *Drosophila* CBP is a co-activator of cubitus interruptus in hedgehog signalling. *Nature*, **386**, 735–738.
21. Waltzer,L. and Bienz,M. (1999) A function of CBP as a transcriptional co-activator during Dpp signalling. *EMBO J*., **18**, 1630–1641.
22. Akimaru,H., Hou,D.X. and Ishii,S. (1997) *Drosophila* CBP is required for dorsal-dependent twist gene expression. *Nature Genet*., **17**, 211–214.
23. Fu,D., Wen,Y. and Ma,J. (2004) The co-activator CREB-binding protein participates in enhancer-dependent activities of Bicoid. *J. Biol. Chem*., **279**, 48725–48733.
24. Zhao,C., York,A., Yang,F., Forsthoefel,D.J., Dave,V., Fu,D., Zhang,D., Corado,M.S., Small,S., Seeger,M.A. and Ma,J. (2002) The activity of the *Drosophila* morphogenetic protein Bicoid is inhibited by a domain located outside its homeodomain. *Development*, **129**, 1669–1680.
25. Zhao,C., Fu,D., Dave,V. and Ma,J. (2003) A composite motif of the *Drosophila* morphogenetic protein bicoid critical to transcription control. *J. Biol. Chem*., **278**, 43901–43909.
26. Zhu,W., Foehr,M., Jaynes,J.B. and Hanes,S.D. (2001) *Drosophila* SAP18, a member of the Sin3/Rpd3 histone deacetylase complex, interacts with Bicoid and inhibits its activity. *Dev. Genes Evol*., **211**, 109–117.

27. Pile,L.A. and Wassarman,D.A. (2000) Chromosomal localization links the SIN3-RPD3 complex to the regulation of chromatin condensation, histone acetylation and gene expression. *EMBO J.*, **19**, 6131–6140.
28. Ludlam,W.H., Taylor,M.H., Tanner,K.G., Denu,J.M., Goodman,R.H. and Smolik,S.M. (2002) The acetyltransferase activity of CBP is required for wingless activation and H4 acetylation in *Drosophila melanogaster*. *Mol. Cell Biol.*, **22**, 3832–3841.
29. Driever,W., Ma,J., Nusslein-Volhard,C. and Ptashne,M. (1989) Rescue of bicoid mutant *Drosophila* embryos by bicoid fusion proteins containing heterologous activating sequences. *Nature*, **342**, 149–154.
30. Sauer,F., Hansen,S.K. and Tjian,R. (1995) DNA template and activator-coactivator requirements for transcriptional synergism by *Drosophila* bicoid. *Science*, **270**, 1825–1828.
31. Sauer,F., Hansen,S.K. and Tjian,R. (1995) Multiple TAFIIs directing synergistic activation of transcription. *Science*, **270**, 1783–1788.
32. Janody,F., Sturny,R., Schaeffer,V., Azou,Y. and Dostatni,N. (2001) Two distinct domains of Bicoid mediate its transcriptional downregulation by the Torso pathway. *Development*, **128**, 2281–2290.
33. Schaeffer,V., Janody,F., Loss,C., Desplan,C. and Wimmer,E.A. (1999) Bicoid functions without its TATA-binding protein-associated factor interaction domains. *Proc. Natl Acad. Sci. USA*, **96**, 4461–4466.
34. Singh,N., Zhu,W. and Hanes,S.D. (2005) Sap18 is required for the maternal gene bicoid to direct anterior patterning in *Drosophila melanogaster*. *Dev. Biol.*, **278**, 242–254.
35. Houchmandzadeh,B., Wieschaus,E. and Leibler,S. (2002) Establishment of developmental precision and proportions in the early *Drosophila* embryo. *Nature*, **415**, 798–802.
36. Spirov,A.V. and Holloway,D.M. (2002) Making the body plan: precision in the genetic hierarchy of Drosophila embryo segmentation. *In Silico. Biol.*, **3**, 89–100.
37. Lucchetta,E.M., Lee,J.H., Fu,L.A., Patel,N.H. and Ismagilov,R.F. (2005) Dynamics of *Drosophila* embryonic patterning network perturbed in space and time using microfluidics. *Nature*, **434**, 1134–1138.

Solution Structure of the K50 Class Homeodomain PITX2 Bound to DNA and Implications for Mutations That Cause Rieger Syndrome[†,‡]

Beth A. Chaney,[§] Kimber Clark-Baldwin,[§] Vrushank Dave,[||,⊥] Jun Ma,[||] and Mark Rance*[,§]

Department of Molecular Genetics, Biochemistry, and Microbiology, College of Medicine, University of Cincinnati, 231 Albert Sabin Way, Medical Sciences Building, Cincinnati, Ohio 45267-0524, Division of Developmental Biology, Cincinnati Children's Hospital Research Foundation, 3333 Burnet Avenue, Cincinnati, Ohio 45229, and Division of Pulmonary Biology, Cincinnati Children's Hospital Research Foundation, 3333 Burnet Avenue, Cincinnati, Ohio 45229

Received December 20, 2004; Revised Manuscript Received March 20, 2005

ABSTRACT: We have determined the solution structure of a complex containing the K50 class homeodomain Pituitary homeobox protein 2 (PITX2) bound to its consensus DNA site (TAATCC). Previous studies have suggested that residue 50 is an important determinant of differential DNA-binding specificity among homeodomains. Although structures of several homeodomain−DNA complexes have been determined, this is the first structure of a native K50 class homeodomain. The only K50 homeodomain structure determined previously is an X-ray crystal structure of an altered specificity mutant, Engrailed Q50K (EnQ50K). Analysis of the NMR structure of the PITX2 homeodomain indicates that the lysine at position 50 makes contacts with two guanines on the antisense strand of the DNA, adjacent to the TAAT core DNA sequence, consistent with the structure of EnQ50K. Our evidence suggests that this side chain may make fluctuating interactions with the DNA, which is complementary to the crystal data for EnQ50K. There are differences in the tertiary structure between the native K50 structure and that of EnQ50K, which may explain differences in affinity and specificity between these proteins. Mutations in the human *PITX2* gene are responsible for Rieger syndrome, an autosomal dominant disorder. Analysis of the residues mutated in Rieger syndrome indicates that many of these residues are involved in DNA binding, while others are involved in formation of the hydrophobic core of the protein. Overall, the role of K50 in homeodomain recognition is further clarified, and the results indicate that native K50 homeodomains may exhibit differences from altered specificity mutants.

The homeodomain is an evolutionarily conserved protein domain found in organisms ranging from yeast and *Drosophila* to humans (*1*−*5*). Homeodomain-containing proteins are known to play important roles in such diverse activities as embryonic pattern formation, cell-type specification, and differentiation (*3*). This domain is responsible for recognizing specific DNA sequences, thereby recruiting the corresponding transcription factors to specific target genes. Most DNA sites that homeodomains recognize consist of only 6 base pairs, with a common TAAT core sequence followed by two base pairs that have been proposed to define the specificity (*3*, *4*). The homeodomain consists of a self-folding, stable protein domain of 60 amino acids. Previous studies of homeodomains have shown that they have a compact 3-helix structure and a flexible N-terminal arm (*3*, *5*). The third helix, called the recognition helix, makes specific contacts within the major groove of the DNA. Homeodomains have been studied extensively (*1*−*5*), including genetic, biochemical, and structural analyses, due to their critical role in cellular processes and to the fact that they serve as a valuable model for probing the physical basis of protein−DNA interactions. Although the overall topology of the homeodomain motif and its docking arrangement on duplex DNA are now generally well-defined, fundamental questions remain, particularly in regard to the role of amino acid side chains in defining the specificity of homeodomain−DNA binding and the nature of the interactions of these side chains with the DNA binding sites. Homeodomains have evolved different DNA specificities in part by altering the amino acid residue at position 50, which can interact with base pairs 5 and 6, and to a lesser extent, base pair 4, in the TAATNN consensus binding site. A previous study has shown that each of 6 different amino acids tested at position 50 confers a different DNA binding specificity (*6*). Tucker-Kellogg et al. (*7*) and others have emphasized the point that the degree of specificity of a homeodomain for its particular DNA binding sites depends on the identity of the amino acid residue in position 50. Most homeodomains contain a glutamine residue at this position, and are therefore referred to as Q50 homeodomains. Q50 homeodomains prefer DNA sequences such as TAATTA and TAATGG. The homeodomain of Bicoid, which is a *Drosophila* morphogenetic protein, contains a lysine at

[†] This work is supported by Grant GM063855 from the National Institutes of Health (to M.R.) and Grant 0255347N from the American Heart Association (to J.M.).

[‡] PDB accession code for PITX2 homeodomain bound to DNA: 1YZ8.

* Author to whom all correspondence should be addressed. Tel: (513) 558-0066. Fax: (513) 558-8474. E-mail: Mark.Rance@UC.Edu.

[§] University of Cincinnati.

[||] Division of Developmental Biology, Cincinnati Children's Hospital Research Foundation.

[⊥] Division of Pulmonary Biology, Cincinnati Children's Hospital Research Foundation (present address).

10.1021/bi0473253 CCC: $30.25
Published on Web 04/26/2005

position 50 and is the founding member of the K50 class of homeodomains (*8*). The K50 class of homeodomains recognizes a consensus DNA sequence of TAATCC. Much attention has been focused on the consequences of lysine being located at position 50, largely due to the fact that the most dramatic examples of altered DNA specificity occur when a lysine is either introduced or replaced at position 50. For example, when Q50 in Engrailed is mutated to an alanine, the Q50A mutant has an affinity and specificity that are very similar to those of the wild-type protein, but when mutated to a lysine, the specificity changes from TAATTA to TAATCC, clearly demonstrating the important role played by the residue in position 50, especially in the case of K50, in defining the specificity of DNA binding (*9*−*11*). Percival-Smith et al. (*12*) investigated wild-type and Q50K mutant Fushi tarazu homeodomains in conjunction with altering the base pairs at positions 5 and 6 in the binding site, and found that differences in K_D of ∼100-fold are observed when the binding site is not the optimal one. In addition to position 50, position 47 has a role in defining specificity for some homeodomains in correlation with base pair 4 of the binding site, especially when the residue is phenylalanine or arginine (*13*, *14*).

Although structures of several homeodomains and homeodomain−DNA complexes have been determined by X-ray crystallography or NMR spectroscopy (*15*), including representatives of the wild-type Q50, S50, C50, G50, and I50 classes of homeodomains (*16*−*20*), the only experimentally determined K50 homeodomain structure available is an X-ray crystal structure of an altered specificity mutant, Engrailed Q50K (EnQ50K[1]), bound to the TAATCC site (*7*). The latter study found that the side chain of K50 projects into the major groove of the DNA and makes hydrogen bond contacts with the O6 and N7 atoms of the guanines at base pairs 5 and 6 of the complementary strand of the TAATCC binding site. This is the only case in which direct hydrogen bond contacts have been reported for amino acid residue 50 in any homeodomain−DNA complex structure. Unfortunately, the relevance of the EnQ50K studies, or analyses of other mutants such as Paired S50K (*6*) and Fushi tarazu Q50K (*21*), to the case of native K50 homeodomains is unclear in the absence of experimental structural data for a native K50 homeodomain. For example, the identity of the amino acid residue at position 54 seems to be constrained by the residue at position 50. A glutamine at position 50 allows for many different residues to be present at position 54, with Met being the most abundant. However, Met54 is never found when position 50 is lysine (*22*). Determining the biological relevance of studies of single site mutants should take into account possible covariation of residues (*23*). For example, structural studies of an EnQ50A mutant have been conducted (*11*) in order to provide additional information concerning the role of residue 50 in general, and Q50 in particular; however, a phage display selection of Engrailed mutants failed to recover a Q50A mutant: the only Q50A mutant recovered also contained a I47T mutation (*24*). Another issue concerning the Engrailed Q50K mutant is the observation that it binds to the consensus TAATCC site with an unusually high affinity, which approaches the picomolar range (*9*). There is no evidence that natural K50 class homeodomains have such a high affinity for DNA (*25*, *26*). The full-length PITX2 protein has a K_D of 50 nM (*25*). A K_D was determined for the Q50K mutant of the Fushi tarazu homeodomain, and this value was found to be 0.63 nM (*12*), which is a much lower affinity than the EnQ50K mutant. The K_D for the PITX2 homeodomain alone was found to be 2.6 ± 0.38 nM (see Supporting Information), which is comparable to the Fushi tarazu mutant, and also a much lower affinity than the Engrailed Q50K mutant. Moreover, the X-ray structure of EnQ50K reveals two conformations with the side chain of K50 contacting either the 5th or 6th position on the antisense strand of the DNA. Whether or not natural K50 homeodomains exhibit these two conformations in a static state and whether or not the side chain of K50 exhibits a fluctuating state were unknown. Another biochemical property shared by the wild-type K50 class homeodomains Bicoid and PITX2 is that they both can recognize naturally occurring DNA sites that deviate from the consensus site TAATCC (*2*, *21*, *27*−*29*). In contrast, a Q50K mutant of the Fushi tarazu homeodomain is unable to bind any nonconsensus DNA binding sites tested (*21*). Together, these considerations underscore the importance of obtaining solution structures of native K50 class homeodomains.

The question regarding side-chain conformational heterogeneity, referred to above in the context of the observations concerning the K50 side chain in the EnQ50K crystal structure, is broader in scope and of fundamental importance for understanding the full range of interactions that can occur at a protein−DNA interface. Crystallographic studies have generally indicated that there are several conserved and relatively stable contacts at the homeodomain−DNA interface. In several instances, such as the aforementioned case of K50 in the EnQ50K structure and the case of Gln50 in the crystal structure of an Even-skipped homeodomain complex (*30*), multiple, significantly populated conformations are observed for the side chain, while the nearly invariant asparagine in position 51 is observed to make very stable contacts with the N6 and N7 atoms of the adenine base in position 3 of the consensus TAAT core binding site. On the other hand, NMR studies (*31*, *32*) and molecular dynamics simulations (*31*, *33*) have provided strong indications of a dynamic, fluctuating environment encompassing some of the key amino acid side chains at the interface, most importantly, the side chains of asparagine 51 and of the position 50 residue. Billeter and co-workers (*31*) proposed that, at least in the case of Antennapedia, the homeodomain achieves specificity through a fluctuating network of short-lived contacts that allow it to recognize DNA without the entropic cost that would result if side chains were immobilized upon DNA binding. Significant interest has been expressed in the literature (*7*, *31*, *33*, *34*) for obtaining experimental data on native K50 homeodomains in order to shed further light on these fundamental issues.

For the structure determination studies reported herein, we chose the well-characterized homeodomain PITX2 as our

[1] Abbreviations: PITX2, pituitary homeobox protein 2; EnQ50K, Engrailed Q50K mutant; IPTG, isopropyl-β-D-thiogalactopyranoside; DSS, 2,2-dimethyl-2-silapentane-5-sulfonic acid; PBS, phosphate buffer solution; TCB, thrombin cleavage buffer; DTT, dithiothreitol; NOESY, nuclear Overhauser enhancement spectroscopy; HSQC, heteronuclear single quantum correlation; NOE, nuclear Overhauser effect; rmsd, root-mean-square deviation; bb, backbone; PDB, Protein Data Bank; HD, homeodomain.

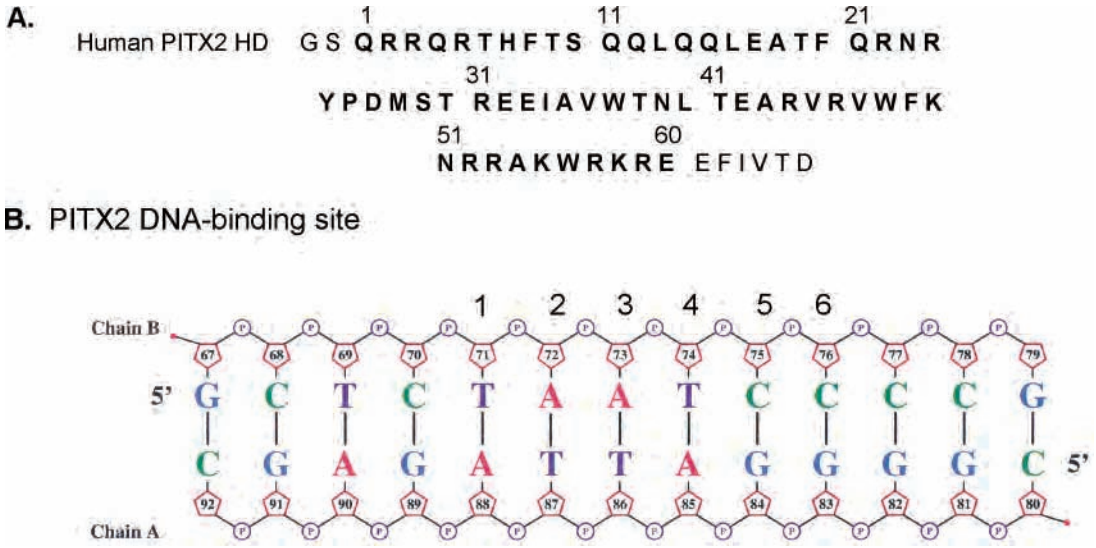

FIGURE 1: (a) Amino acid sequence of the PITX2 homeodomain used for the structural studies. The extra residues at the N- and C-termini are part of the expression system. (b) DNA sequence of the binding site used in the structural studies. The DNA sequence consists of the TAATCC binding site surrounded by residues to confer stability to the double strand.

representative model for the K50 class of homeodomains. PITX2 is a transcription factor that is found in many developing tissues in vertebrate embryos. It has been shown to be expressed in the brain, heart, pituitary, mandibular and maxillary regions, eye, gut, limbs, and umbilicus (*35*−*38*). The homeodomain is just one domain of the full-length PITX2 protein. There have been three major isoforms of PITX2 identified, and these isoforms are produced by alternative splicing and the use of different promoters (*35*, *36*, *39*−*41*). All of the isoforms contain different N-terminal domains, while the homeodomain and C-terminal domains are identical. The C-terminal region contains a transcriptional activation domain. Mice that have been genetically engineered to be homozygous for a *pitx2* null allele have a single atrium, arrested development of the pituitary gland, numerous defects of the eye, and altered development of the mandibular and maxillary regions (*41*−*43*). Mutations in *PITX2* are known to be a cause of Rieger syndrome (*44*−*48*), a genetic disease characterized by defects in the eye, facial features, and umbilicus. Sequencing of DNA from human patients has shown that many mutations in *PITX2* result in single amino acid substitutions within the homeodomain region (*45*−*48*). Although modeling and biochemical studies have provided useful insights into the effect of these mutations (refer to Table 2), no structure-based information is available on how these mutations might affect the PITX2 homeodomain properties.

In the present study, the NMR solution structure of the PITX2 homeodomain bound to its consensus DNA binding site is reported. This represents the first experimentally determined structure of a native K50 class homeodomain. The results reveal a tertiary structure similar to other homeodomains, with K50 making contacts with the two guanines adjacent to the TAAT core DNA sequence. Evidence indicates that K50 may interact with DNA in a flexible manner. The tertiary structure of PITX2 indicates that the first two helices are slightly closer to each other than in the EnQ50K mutant, and the third helix has a slightly different position relative to the other helices. We discuss how these structural properties of PITX2 might affect the biochemical functions of the PITX2 homeodomain. On the basis of the solution structure, we also discuss the effect of mutations found in Rieger syndrome patients on PITX2 homeodomain functions.

MATERIALS AND METHODS

Expression of the PITX2 Homeodomain. The expression plasmid pGEX-1λt-PITX2HD consists of a glutathione S-transferase tag and a thrombin cleavage site prior to the PITX2 homeodomain sequence. There are two extra residues at the N-terminus as a result of thrombin cleavage, and six extra residues at the C-terminus that are part of the expression system. The final protein sequence is shown in Figure 1.

The PITX2 homeodomain was obtained by growing *Escherichia coli* strain BL21-Star (Invitrogen) transformed with pGEX-1λt-PITX2HD in minimal medium [0.85 g/L NaOH, 10.5 g/L K_2HPO_4, 12 g/L Na_2HPO_4, 6 g/L KH_2PO_4, 1 g/L NaCl, 6 mg/L $CaCl_2$, 13.2 mL/L concentrated (12.2 N) HCl, nucleotides (0.5 g/L adenine, 0.65 g/L guanosine, 0.2 g/L thymine, 0.5 g/L uracil, 0.2 g/L cytosine), vitamins (1 mg/L choline chloride, 1 mg/L pyridoxal phosphate, 100 μg/L riboflavin, 50 mg/L thiamine, 50 mg/L niacin, 1 mg/L biotin), and trace elements (107 μg/L $MgCl_2 \cdot 6H_2O$, 20 μg/L $FeCl_2 \cdot 4H_2O$, 0.7 μg/L $CaCl_2 \cdot 2H_2O$, 0.26 μg/L H_3BO_3, 0.16 μg/L $MnCl_2 \cdot 4H_2O$, 16 ng/L $CuCl_2 \cdot 2H_2O$, 2.4 μg/L $Na_2MoO_4 \cdot 2H_2O$, 10 μM $FeCl_3$, 135 mM $CaCl_2$, 50 μM $ZnSO_4$)] containing 150 mg/L ampicillin, 4 g/L glucose, and 1 g/L NH_4Cl. Half a liter of bacterial culture was grown in baffled flasks in an incubator shaker at 37 °C until saturation (A_{600} $\sim$ 5.0). This culture was spun down (2000*g*, 10 min) and resuspended in 1 L of minimal medium enriched with 10% $^{15}N^{13}C$-Isogro, ^{13}C-glucose, and ^{15}N-ammonium chloride (Sigma-Aldrich). Expression was then induced by addition of IPTG to a final concentration of 0.1 mM, followed by growth at 20 °C for approximately 24 h.

Purification of the PITX2 Homeodomain. Cells were lysed with lysozyme and sonication. Cleared lysate was applied to a glutathione sepharose column and washed with PBS buffer, followed by thrombin cleavage buffer (50 mM trizma hydrochloride, 150 mM NaCl, 2.5 mM $CaCl_2$, pH 8.0). The resin was resuspended in thrombin cleavage buffer (TCB)

and transferred to a 50 mL conical tube. The homeodomain was cleaved from the glutathione S-transferase fusion tag using 1 mg of thrombin for 3 h at 4 °C. Nearly complete cleavage was obtained during this time as measured by SDS−PAGE. The cleaved protein was then eluted from the resin using 5 bed volumes of TCB. It was loaded onto a SP sepharose fast flow column (2 mL bed volume, Amersham), washed with washing buffer (10 mM NaH_2PO_4, 250 mM NaCl, pH 7.0), and eluted with buffer containing a higher salt concentration (10 mM NaH_2PO_4, 1 M NaCl, pH 7.0). Fractions containing the homeodomain were identified by Abs_{278} (extinction coefficient = 18350 cm^{-1} M^{-1}), pooled, and dialyzed overnight at 4 °C in 10 mM NaH_2PO_4, pH 7.0. Protein yields were ~4.5 mg/L of cell growth. The consensus DNA duplex (IdtDNA) (see Figure 1) was added to give a 1:1 protein:DNA ratio, and the complex was concentrated by burying the dialysis bag in Spectra/Gel Absorbent (Spectrum), or Aquacide (Calbiochem). Samples were dialyzed in 10 mM NaH_2PO_4, pH 7.0 after concentration. Complete protease inhibitors (Roche, 1 tablet in 3 mL, add 1 μL), leupeptin (0.3 mM), DTT (2 mM), and Pefabloc (0.2 mM, Roche) were all added to inhibit proteases. Sodium azide (6 mM stock, add 1 μL to 540 μL NMR sample) was added to prevent bacterial growth in the sample. The final sample concentration was approximately 1 mM, in 90% 10 mM NaH_2PO_4, pH 7.0 and 10% D_2O.

Determination of K_D. Measurements of K_D were taken by following the procedure of Dave et al. (*2*). The DNA probe concentrations used in this analysis were 1, 2, 4, 8,16, 20, and 40 nM. Quantitative gel shift assays were performed by measuring the bound and free fractions of the probes with a PhosphorImager as previously described (*21*). The data were analyzed using Microsoft Excel (linear regression analysis) to determine the K_D value ($-1/K_D$ = slope of the plot of bound/free against bound DNA). These results can be seen in the Supporting Information.

NMR Structure Determination. All NMR experiments were carried out on Varian Inova 600 and 800 MHz spectrometers. The sample temperature was set to 295 K. Spectra were referenced to an external DSS standard.

Protein Assignments. Protein ^1H, ^{13}C, and ^{15}N resonance assignments were obtained primarily from heteronuclear-edited NMR spectra, using conventional triple resonance ^1H-$\{^{13}C,^{15}N\}$ NMR probes. The pulse programming codes were written in-house. Approximately 92% of assignable atoms were assigned. Sequence-specific assignment of the backbone H^N, N, C′, C^α, and C^β resonances were obtained from 3D HNCO, HN(CO)CA, HNCA, CBCA(CO)NH, and HNCACB (*49*−*55*) spectra. Assignment of the aliphatic side-chain resonances was accomplished using a combination of 3D ^{15}N-edited-TOCSY-HSQC (*56*, *57*), H(CCO)NH-TOCSY, and HBHA(CBCACO)NH spectra (*55*). Aromatic ^1H and ^{13}C resonances were obtained from a combination of 2D HMQC, 2D HMQC-TOCSY, 3D HMQC-TOCSY, and 2D NOESY-HMQC spectra (*58*, *59*). A HNHA experiment was performed to assign H^α and to obtain coupling constants (*60*).

DNA Assignments. Resonance assignments for unlabeled DNA bound to ^{13}C,^{15}N-labeled protein were obtained using standard assignment methods for DNA (*61*). The data was obtained with doubly $^{13}C/^{15}N$-filtered NOESY and ω_2-filtered TOCSY experiments (*62*, *63*).

Structural Constraints. The main source of structural information was the proton−proton distance constraints identified from NOESY spectra. Three-dimensional ^{15}N-NOESY-HSQC experiments (*64*) using 50−125 ms mixing times were used for intramolecular restraints in the homeodomains, along with a ^{13}C-NOESY-HSQC experiment using a 150 ms mixing time (*55*).

Intramolecular distance restraints for the DNA were obtained from an r^{-6} scaling of cross-peak volumes in the NOESY spectra. Upper and lower bounds were calibrated on the cytosine intraresidue H5−H6 NOE and set to ± 15% of the calculated distance for base and H1′ protons. Restraint boundaries to other sugar protons were widened an additional 10% to account for effects of spin diffusion. Restraints from the longer mixing time 125 ms experiment were assigned a lower bound of 3 Å and an upper bound of 5 Å.

Intermolecular restraints between the protein and DNA were obtained from 2D $^{13}C(\omega_1)$-edited, $[^{13}C,^{15}N](\omega_2)$-filtered NOESY spectra (*63*, *65*, *66*). The NOEs were assigned manually, and only unambiguously assigned peaks were used as restraints in the docking calculation. Weak peaks were assigned an upper distance limit of 6.0 Å, while medium peaks had an upper distance limit of 5.0 Å, and stronger peaks an upper distance limit of 4.0 Å.

Data Processing and Analysis. Raw NMR data was processed using NMRPipe (*67*). Linear prediction was used in the t_1 dimension for 2D spectra and in the t_1 and t_2 dimensions for 3D spectra, using sinebell window functions for apodization and zero filling in all dimensions. Spectra were viewed and analyzed using the Sparky graphical interface (*68*). This program was used to pick peaks and integrate them using a Lorentzian function.

Structure Calculation. Referenced chemical shift assignments and peak intensities from Sparky were entered into the structure calculation program CYANA (*69*, *70*). CYANA consists of an automated NOE assignment program, CANDID, which automatically assigns all NOESY cross-peaks, taking into account nearness of chemical shift, network anchoring, ambiguous distance constraints, and constraint combination (*70*). The structure is then calculated using the DYANA algorithm, which calculates structures using torsion angle dynamics (*69*). Calibration constants for peak intensities versus upper distance limits were determined automatically by CYANA. The 20 lowest energy conformers were retained after structure calculation and used for docking to DNA.

Docking of the Protein to the DNA. The protein was docked to the DNA using the AMBER all-atom force field with the generalized Born solvation model (*71*, *72*). The 20 CYANA structures with the lowest values of target function were docked onto canonical B-form DNA. This was chosen as the starting DNA structure, since NOESY spectra for the complex indicated the DNA to be close to B-form. Starting structures of the complex were generated by systematically placing PITX2 in varying orientations relative to the DNA, with helix 3 approximately 50 Å from the DNA. For each of the 20 lowest-energy structures, 5 different orientations of the protein relative to the DNA were selected, yielding 100 starting conformers.

The protein was docked onto the DNA by a 20 ps simulated annealing calculation (T = 600 K, time step = 1 fs) using an altered version of a procedure described

previously for docking of TFIIIA to DNA (*73*). The temperature was increased from 0 to 600 K over the first 4 ps, held at 600 K for 2 ps, and then slowly cooled to 0 K over 14 ps. The weights of the force constants were linearly increased from 0.1 to 1 during the course of the calculation. DNA base-pairing was maintained by incorporating Watson−Crick hydrogen-bonding restraints. These Watson−Crick DNA restraints were implemented as lower and upper bound restraints on base-paired heteroatom−heteroatom (2.7 to 3.1 Å) and heteroatom−proton distances (1.67 to 2.07 Å), and had a final force constant of 50 kcal mol^{-1} $Å^{-2}$. The intramolecular protein and DNA restraints had final force constants of 20 kcal mol^{-1} $Å^{-2}$. Protein and DNA angle restraints had a final force constant of 32 kcal mol^{-1} $Å^{-2}$. Protein−DNA intermolecular restraints had a final force constant of 32 kcal mol^{-1} $Å^{-2}$. Protein restraints were applied to prevent the protein conformation from being altered too much from the structure calculated by CYANA. DNA restraints were applied to prevent fraying of the DNA, and to maintain the structure close to B-form.

Structure Refinement and Analysis. The 20 structures with the lowest AMBER energy values were subjected to restrained energy minimization by the SANDER module of the AMBER 7.0 package (*71*, *72*). The 1994 version of the force field was used. Each conformer was subjected to a conjugate-gradient energy minimization calculation with solvent included.

The evaluation of the structure, i.e., analysis of geometry, stereochemistry, and energy distributions in the models, was performed using the program PROCHECK (*74*). Restraint violations were analyzed using the program AQUA (*75*). Graphics were prepared using MOLMOL (*76*).

RESULTS AND DISCUSSION

Structure Determination. Assignments of the protein backbone and side-chain ^{1}H, ^{13}C, and ^{15}N resonances were obtained from heteronuclear spectra. Restraint data derived for the PITX2 homeodomain−DNA complex are summarized in Table 1. Analysis of ^{15}N and ^{13}C heteronuclear-edited NOESY spectra recorded at various mixing times provided 1259 intramolecular distance restraints comprising 513 intraresidue, 338 sequential, 300 medium-range, and 108 long-range NOE contacts. Torsional restraints for 55 ϕ and 43 ψ angles were obtained from a 3D HNHA experiment, and from using C^{α} chemical shifts (*77*, *78*). Overall, there are 19 restraints per residue, on average, for intramolecular protein NOEs. All of these protein restraints were used for structure calculation with the program CYANA (version 1.0.6) (*69*, *70*). After the final round of structure calculation, the 20 structures with the lowest CYANA target function were used for docking to DNA and energy minimization. The final average CYANA target function for the 20 structures was 2.05 $Å^2$.

A total of 292 distance restraints between protons within the DNA were obtained from $^{13}C/^{15}N$-filtered NOE spectra. A series of 2D $^{13}C(\omega_1)$-edited, $[^{13}C,^{15}N](\omega_2)$-filtered NOESY spectra provided 27 unambiguous intermolecular restraints between the protein and the DNA. These restraints were entered into the program AMBER for docking and energy minimization, as described in Materials and Methods.

Quality of the NMR Structure. The structure of the PITX2−DNA complex was calculated by a restrained molecular dynamics docking and energy minimization procedure starting from the coordinates of the PITX2 protein calculated from CYANA and canonical B-form DNA as described in Materials and Methods. The 20 structures with the lowest total energies were selected for conformer analysis. These structures exhibited mean AMBER energies of −6268 kcal mol^{-1} and mean van der Waals and electrostatic energies of −399 and −3974 kcal mol^{-1}, respectively. The mean AMBER energies given represent the intraprotein interaction energy.

Table 1. NMR Structure Statistics[a]

NMR constraints	
protein	
distance constraints	1259
intraresidue	513
sequential	338
medium-range	300
long-range	108
dihedral constraints	98
phi	55
psi	43
DNA	292
protein−DNA (intermolecular)	27
total	1676
CYANA target function value ($Å^2$)[b]	2.05 ± 0.39
no. of violations	
distance violations (>0.30 Å)	0
dihedral angle violations (>5.0°)	1
AMBER energies (kcal/mol)[c]	
mean AMBER energy	−6268 ± 250
van der Waals	−399 ± 33
electrostatic	−3974 ± 336
Ramachandran plot (%)[d]	
residues in most favored regions	80.1
residues in additional allowed regions	14.8
residues in generously allowed regions	2.3
residues in disallowed regions	2.8
rmsd from the mean structure (Å)	
protein (bb, residues 3−58)	1.38
all heavy atoms (residues 3−58)	1.95
protein (bb, all residues)	1.85
DNA (residues 68−78, 81−91)	1.30
complex (residues 3−58, 68−78, 81−91)	1.81

[a] A total of 30 conformers were calculated, and the 20 structures with the smallest residual CYANA target function values were subjected to docking and energy minimization. [b] The value given for the CYANA target function corresponds to the value before energy minimization (the CYANA target function is not defined after energy minimization, since the conformers no longer have ECEPP standard geometry). [c] The value given represents the intra-protein interaction energy. [d] For residues 3−58.

The superposition of the structures (Figure 2) demonstrates a well-defined tertiary structure for PITX2 bound to DNA. The structures have no distance violations greater than 0.3 Å, and only 1 angle violation greater than 5°. Analysis of Ramachandran plots for the ensemble indicates that the structures generally show favorable backbone conformations within allowed conformational space, with 80.1% of the residues 3−58 within the most favored regions, 14.8% in additionally allowed regions, 2.3% in generously allowed regions, and 2.8% in disallowed regions for the 20 conformers (Table 1). The N- and C-termini are largely disordered. When superimposed, residues 3−58 have an average root-mean-square deviation (rmsd) from the mean structure of 1.38 Å for backbone (N, C^{α}, C′, and O), 1.95 Å for all heavy

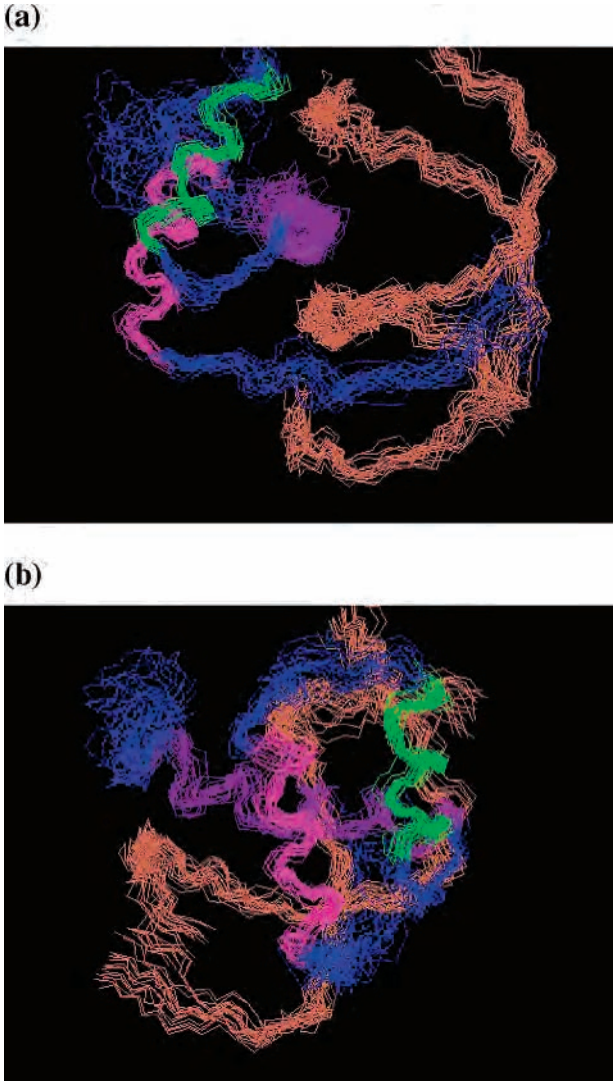

FIGURE 2: Ensemble of structures of the PITX2 homeodomain−DNA complex. (a) Ensemble of 20 structures showing the protein backbone N, C^{α}, and C′ atoms and the DNA backbone. Helix 1 is colored pink, helix 2 green, helix 3 purple, and the DNA strands are coral. Superimposition was performed using backbone atoms from protein and DNA. (b) Alternate view of the structure, rotated by approximately 90°.

atoms, and 1.85 Å for the backbone when all residues are included. The global rmsd for all DNA heavy atoms (nucleotides 68−78, 81−91) is 1.30 Å. The rmsd for the entire complex (residues 3−58; nucleotides 68−78, 81−91) is 1.81 Å.

Tertiary Structure of the PITX2 Homeodomain−DNA Complex. The overall tertiary structure of the PITX2 homeodomain is similar to other homeodomains, supporting previous findings that this tertiary structure is well conserved among homeodomains (*1*, *3*, *79*, *80*). The tertiary structure of the PITX2 homeodomain is composed of three α helices (Figure 3). Helix 1 (residues 10−20) is followed by a loop region, and then helix 2 (residues 28−37) runs antiparallel to helix 1. Helix 2 and helix 3 (residues 42−58) form a helix−turn−helix motif. Helix 3 is approximately perpendicular to helices 1 and 2, and fits into the major groove of the DNA. The N-terminus of the homeodomain makes contacts within the minor groove of the DNA.

The helices of the PITX2 homeodomain are held together by a core of eight tightly packed hydrophobic amino acids

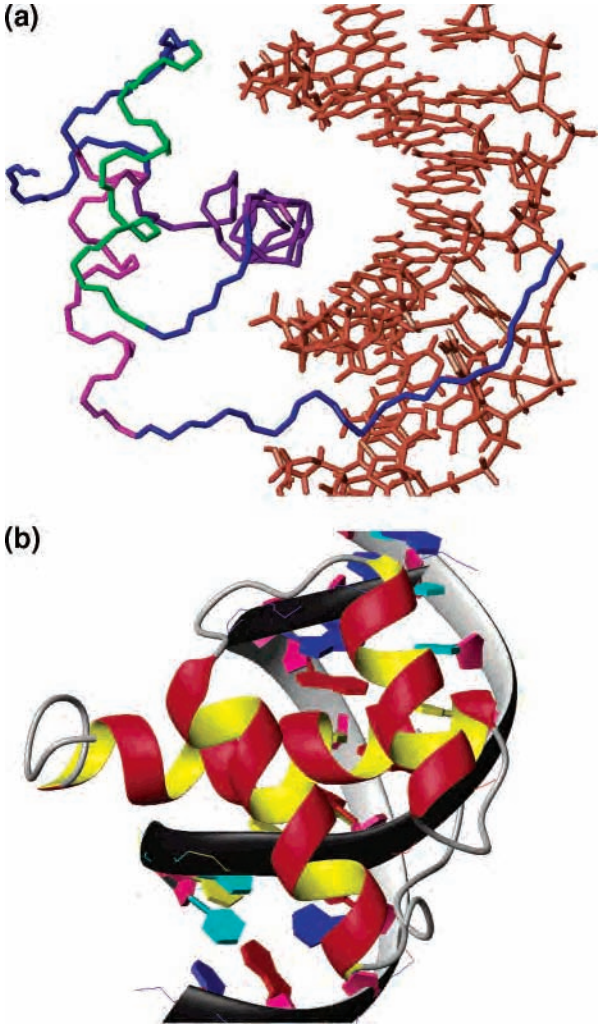

FIGURE 3: Structure of the PITX2 homeodomain−DNA complex. (a) Mean structure of the homeodomain. Helix 1 is colored pink, helix 2 green, and helix 3 purple. The DNA strands are colored coral. (b) Ribbon diagram of the mean structure of the PITX2 homeodomain−DNA complex.

(F8, L13, L16, F20, L40, V45, W48, and F49). These amino acids are either invariant (W48 and F49) or highly conserved in all homeodomains (*3*, *81*, *82*). In a threading analysis performed previously for the PITX2 homeodomain (*83*), it was hypothesized that the tertiary structure of the PITX2 homeodomain would be similar to other homeodomains, mainly because many of these hydrophobic amino acids that are present in other homeodomains are also present in the PITX2 homeodomain. The threading analysis threaded the PITX2 homeodomain sequence to the Engrailed homeodomain structure, so the overall tertiary structure ended up being very close to that of Engrailed. The threading analysis did not provide a PDB file that we could analyze in detail, and in any case is not necessarily indicative of the true molecular structure. For this threading analysis, the focus was the role of Rieger mutations in causing disease, and there was no discussion of the role of K50 in determining the DNA-binding affinity and specificity of the homeodomain, which is something best addressed via an experimentally determined structure rather than a threading model. This study also did not analyze the K50 Rieger mutants, so there is no indication what the structure of this side chain was in their analysis. In the absence of an experimentally determined

structure, the threading model was most useful for visualizing some of the intramolecular interactions that stabilize the tertiary structure, and the predicted interactions are consistent with our experimental data. While we cannot compare our PITX2 tertiary structure directly to that of the threaded structure, we can compare it to EnQ50K and other homeodomain structures. In our experimentally determined structure, the first helix is closer to the second helix when measured from the backbone nitrogen of L16 to the backbone nitrogen of I34 and compared to the EnQ50K, Antennapedia, wild-type Engrailed, Fushi tarazu, vnd/NK-2, and MATα2 homeodomains (*7*, *84*−*88*). As far as this distance is concerned, PITX2 is an outlier compared to the other six homeodomains. This distance is a range of 9.60−10.70 Å for Antennapedia conformers, 9.43 Å for the crystal EnQ50K structure, 9.54 Å for wild-type Engrailed, 9.30−11.10 Å for Fushi tarazu, 8.67 Å for vnd/NK-2, and 10.9 Å for MATα2. However, for PITX2 this distance range over the 20 conformers is only 7.55−8.58 Å, which is an average of 1.8 Å closer. In view of the rmsd for the PITX2 protein backbone atoms (residues 3−58) of 1.38, this result is still significant, when compared with the ranges of distances seen in the structures of the other homeodomains. This difference is especially significant when considering that the rmsd for helices 1 and 2 alone is only 0.78 Å. The range for the distance between L16 and I34 for all of the other homeodomains together is 8.67−11.10 Å, and the PITX2 distance range is completely outside of this.

In addition to the narrower distance between the first and second helices in the PITX2 structure, there are several other differences between the PITX2 and EnQ50K structures. In particular, the third helix of PITX2 is positioned about 0.5 Å lower (closer to the N-terminus of helix 1 and C-terminus of helix 2) than in EnQ50K (Figure 4). This difference in orientation of the three helices causes slightly different contacts to be made between the first and third helices, and may provide a partial explanation for the decreased stability of this homeodomain. Unlike other homeodomains that are stable in the free form (*32*, *86*, *89*−*92*), the PITX2 homeodomain is unstable in the absence of DNA in that it irreversibly aggregates at micromolar concentrations, which suggests a possible lack of stable tertiary structure in the free form. This may be due to slightly different hydrophobic interactions within the core of the protein, and the absence of other stabilizing interactions such as the salt bridge linking residues 19 and 30, which can be present in most homeodomains (*93*) but is not possible in PITX2. One difference seen here is that F49, which is nearly invariant among homeodomains, points slightly upward toward the loop region of the homeodomain, instead of pointing toward the interior of the protein. The orientation of the first helix in relation to the third would cause a steric clash with F49 if it were in an orientation similar to that of other homeodomains. While there is still an interaction involving F49 and F20 within the hydrophobic core of the PITX2 homeodomain, the orientations of the side chains themselves are different. This differing orientation may lessen the strength of the interaction between the first and third helices, which may affect the stability of the protein in the absence of DNA. This difference in orientation may be due to any number of differing residues between the two homeodomains (see Supporting Information). One possibility is a proline residue that is found in the loop region between helices 1 and 2 in PITX2, but is not present in Engrailed or Fushi tarazu.

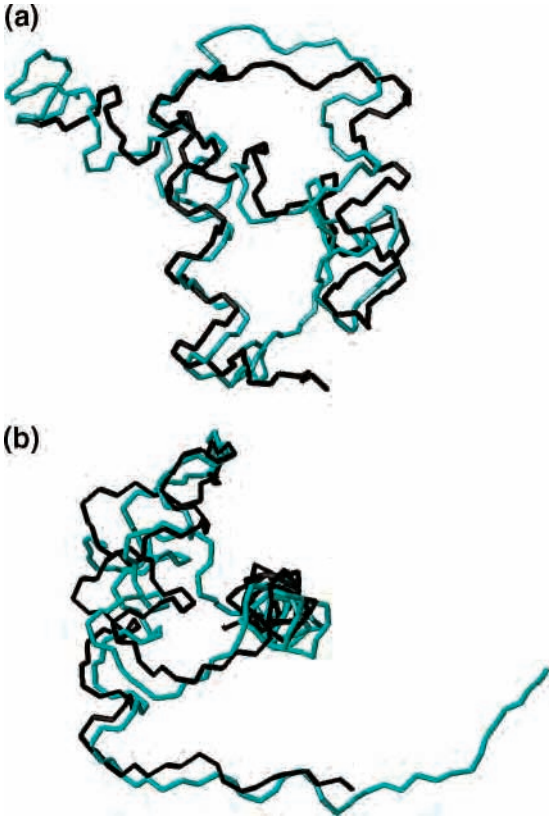

FIGURE 4: Overlay of PITX2 homeodomain and EnQ50K homeodomain structures. Cyan corresponds the structure of the PITX2 homeodomain, and black corresponds to the structure of the Engrailed mutant homeodomain. (a) Helices 1 and 2 are approximately 1.8 Å closer to each other in PITX2 than in other homeodomains. (b) Alternate view, rotated by approximately 90°. Helix 3 is about 0.5 Å lower in PITX2 than in EnQ50K.

In the PITX2 structure, the N- and C-terminal segments −2 to 2 and 60 to 68 (Figure 1) appear disordered (Figure 2), which is to be expected on the basis of a lack of medium-range and long-range constraints for these residues. Analysis of ^{15}N relaxation data (unpublished) indicates that residues −2 to 2 and 59 to 68 are more mobile in solution, explaining the observed disorder and lack of restraint information for these regions.

Our study also reveals structural information about the DNA when it is bound to the protein. Distance restraints obtained from the experiments described above for assigning the DNA were entered into AMBER during the docking procedure. Visual inspection of the structure of the PITX2 homeodomain−DNA complex indicates that there is a slight widening of the minor groove of the DNA compared to B-form DNA, and a concomitant narrowing of the major groove. Previous structures of protein−DNA complexes have indicated that changes in DNA structure are possible upon protein binding (*94*). A more thorough, quantitative analysis of the DNA structure when PITX2 is bound will not be possible until a high-resolution structure of the DNA is determined, using isotopically labeled DNA (*95*).

Protein−DNA Recognition. Analysis of the filtered NOESY experiments produced 27 unambiguous distance restraints between the protein and the DNA (see Supporting Information for a list). These include contacts that have been seen

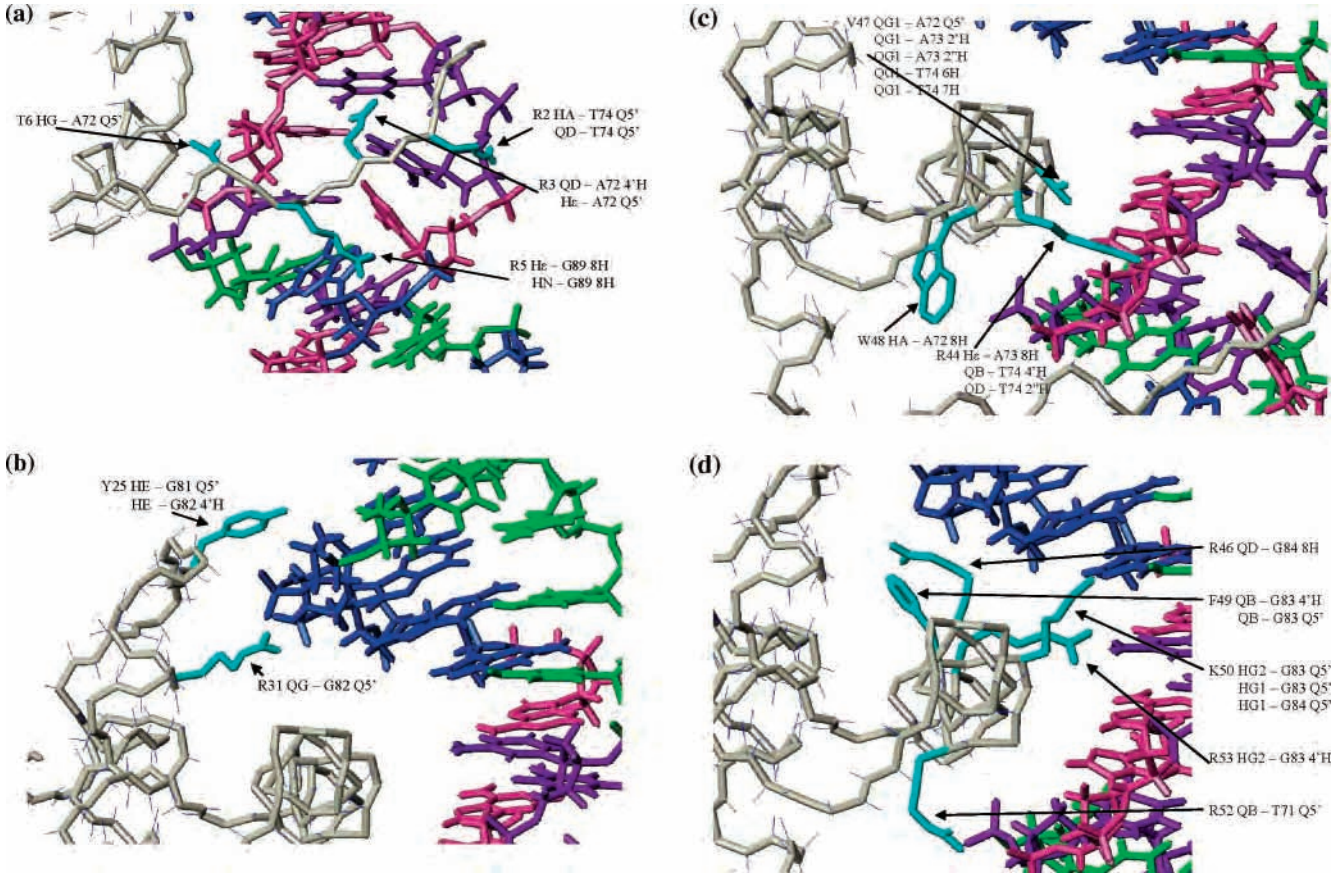

FIGURE 5: Detailed view of the protein−DNA interface and protein−DNA contacts. The backbone of the protein is shown in beige, with the C^{α}−H^{α} bonds shown with small blue lines. Side chains of the protein are illustrated in cyan. On the DNA, blue corresponds to guanine residues, green to cytosine, pink to adenine, and purple to thymine. (a) View of the protein−DNA NOE contacts between the N-terminus of the PITX2 homeodomain and the minor groove of the DNA. (b) View of the protein−DNA NOEs between Y25, R31, and the DNA. (c, d) View of protein−DNA NOE contacts between residues in the third helix and the major groove of the DNA.

in other biochemical and structural studies of homeodomains. Many of the residues that interact with the DNA are arginines, including R3 and R5 at the N-terminus, R31 in the second helix, and R46, R52, and R53 in the third helix (Figure 5). Other residues that were found to make DNA contacts are Y25 and F49. A number of NOESY peaks were also seen between K50 and the DNA, and these contacts are discussed further below.

A detailed picture of the protein−DNA interface is shown in Figure 5. This figure illustrates the orientations of some of the side chains that are important in DNA binding, particularly within the third helix (the specific atom contacts are indicated in Figure 5). Figure 1b outlines the numbering of the DNA used in the following discussion. Figure 5a illustrates the protein−DNA NOE contacts seen within the N-terminal arm. NOE contacts were seen in the minor groove between R2, R3, and R5 and DNA residues A72, T74, and G89. Although the NOESY-derived distance constraints indicate contact between residues R3 and R5 and the minor groove, ^{15}N relaxation data (unpublished) indicates that this region of the N-terminus does retain some degree of mobility; similar results were reported for the Even-skipped homeodomain, on the basis of refined atomic B factors (*30*). Broad line widths were observed for the backbone NH resonances of His7 and Phe8, which are indicative of slow time scale motions in this region of the homeodomain and could possibly render undetectable possible NOEs from these residues to the DNA.

In the second helix, R31 has a NOE contact from the H^{γ} position to G82 Q5′ (Q refers to a pseudoatom representation, to indicate that stereospecific assignment of the 5′ and 5″ protons was not made), as can be seen in Figure 5b. HBPLUS analysis (*96*) indicates that R31 is making a hydrogen bond contact with the phosphate backbone of this nucleotide. In the loop between helices 1 and 2, Y25 H^{ϵ} is making NOE contacts with G81 Q5′ and G82 4′H. In the third helix, V47 $Q^{\gamma 1}$ is making conserved NOE contacts to A72 Q5′, A73 2′H and 2″H, and T74 6H. Residue W48 has a NOE contact between H^{α} and A72 8H. R44 H^{ϵ} is making contact with DNA proton A73 8H, R44 Q^{β} with T74 4′H, and R44 Q^{δ} with T74 2″H. HBPLUS analysis indicates that R44 is making a backbone hydrogen bond contact to the phosphate of T74. Residues 44, 47, and 48 are illustrated in Figure 5c. In the third helix, R46 and R52 appear to be making conserved contacts with the DNA backbone. R46 extends upward, and R52 extends downward to make these contacts (Figure 5d). R46 Q^{δ} has a NOE contact with G84 8H. R52 Q^{β} has a NOE contact with T71 Q5′. R53 Q^{γ} makes a NOE contact with G83 4′H. All of these NOEs could be due to the close proximity of the atoms while the side chains form hydrogen bonds with backbone phosphate groups. NOEs are also seen between F49 Q^{β} and G83 4′H and Q5′. K50 will be discussed further below, but as can be seen in Figure 5d, there are NOE contacts between the K50 side chain and atoms from G83 and G84.

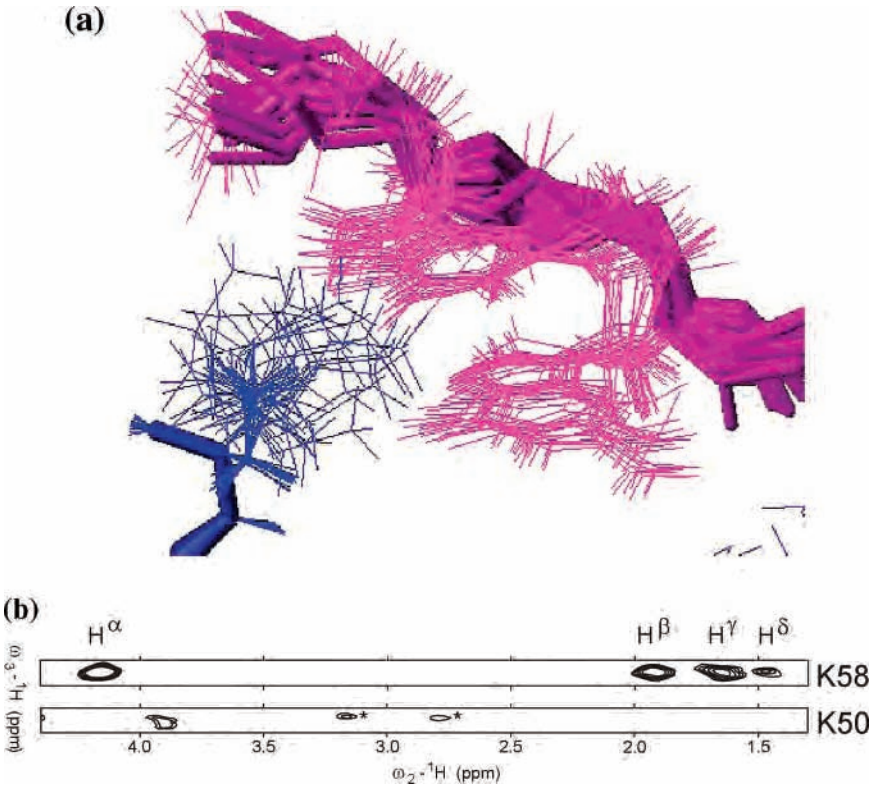

FIGURE 6: (a) View of the 20 conformers, with only the K50, G83, and G84 backbone and side-chain atoms shown to illustrate the extent of disorder of the K50 side chain, implying possible mobility of this side chain in interacting with the DNA. K50 atoms are shown in blue, and G83 and G84 atoms are shown in pink. Backbone atoms are bolder than side-chain atoms. (b) Strips from an H(CCO)NH-TOCSY spectrum showing proton resonances for the side chains of K58 and K50. Line broadening of resonances in the K50 side chain, leading to the weak or missing signals, is indicative of possible motion of this side chain. The asterisks in the K50 slice indicate peaks breaking through from an adjacent ^{15}N plane. The sole K50 peak is for the H^{α} proton resonance.

Other residues that were observed in the calculated structures to be in close contact with the DNA, but without NOEs being seen in the NMR data, are N51, K55, R57, and K58. N51 is nearly invariant among homeodomains (*82*) and is found herein to make the same highly conserved interaction within the major groove with base A73. This residue has been shown in crystal structures to form a pair of hydrogen bonds with this adenine at the N7 and N6 positions, while NMR studies have indicated possible rapidly interchanging conformations (*32*, *97*). NMR studies have shown this close interaction, but no NOEs are seen, possibly due to line-broadening effects (*97*). While no NOEs are seen between K55, R57, or K58 and the DNA, HBPLUS analysis of the complex indicates that there are possible interactions present. K55 may be forming a hydrogen bond with the phosphate of T71. R57 may be contacting the phosphate of G84. K58 may be contacting the phosphate of C70. Due to the usual sensitivity limitations in the edited/filtered NMR experiments employed to identify intermolecular NOEs, it is quite likely that a number of anticipated NOEs fall at or below the threshold for detection.

The Role of Lysine at Position 50. No previous structures have been described for any native K50 class homeodomains. However, the X-ray crystal structure of the Q50K mutant of the Engrailed homeodomain bound to DNA has been reported (*7*), and the side chain of K50 was found to project into the major groove of the DNA, making hydrogen bond contacts with the O6 and N7 atoms of the guanines at base pairs 5 and 6 of the complementary strand of the TAATCC binding site. Our structure of the PITX2 homeodomain marks the first experimentally determined structure of a native K50 class homeodomain, and is important for validating results seen in the studies of non-native proteins. When binding to the consensus site, the position of K50 is very similar to that seen in the EnQ50K structure, with the side chain of K50 extending outward and making contacts with the two guanines adjacent to the TAAT core sequence on the antisense strand (Figure 5d). NOEs are observed between the K50 Q^{γ} and the Q5′ protons of G83 and G84. The N^{ζ} of the K50 side chain is likely making hydrogen bond contacts to the O6 and N7 atoms of G83 and G84, according to analysis by HBPLUS (*96*).

NMR spectroscopy allows one to obtain information about the mobility of the protein backbone and side chains. A key finding in the present study was that the side chain of K50 potentially mediates recognition by fluctuating between multiple conformations. The conformational heterogeneity can be seen in Figure 6a. This preliminary evidence is based on averaging of NOEs and broadening of resonances for this residue. The averaging of NOEs was dealt with as ambiguous distance constraints within the structure calculation in CYANA, and these constraints were satisfied in all structures of the family. When results from an H(CCO)NH-TOCSY experiment are compared between the K50 and K58 side chains (Figure 6b), peaks are easily seen for the K58 side-chain resonances, but only the H^{α} resonance is seen for the K50 side chain. The extra peaks in the K50 strip of Figure 6b are from another residue on an adjacent nitrogen plane and are strong enough to show up as residual peaks on this plane. The broadening of resonances for this side chain made

it difficult to assign using typical heteronuclear-edited NMR spectra. Instead, assignments were made using NOESY spectra and eliminating assignments from nearby residues, until only K50 resonances were left. In principle, it is possible that the line broadening of K50 side-chain resonances could be caused by ring current effects from aromatic bases in the DNA, or by mobility of other nearby protons in the DNA binding site. However, no anomalous line broadening was observed for DNA proton resonances in the vicinity of the K50 side chain. In addition, results similar to those reported here have been seen in other DNA-binding proteins in which side-chain mobility appears to cause line broadening of resonances (vide infra) (*32*, *97*−*99*). These results, in combination with the multiple conformations observed for K50 in EnQ50K, provide compelling evidence that the side chain of K50 is mobile. Preliminary ^{15}N relaxation measurements of the homeodomain backbone dynamics (unpublished) did not show anything unusual in the region of K50. Some degree of side-chain mobility at the protein−DNA interface would be expected to confer an entropic advantage for binding to the DNA. It has been estimated previously that the entropic cost of keeping a lysine side chain static during binding is 3 kcal mol^{-1} (*100*). This possible entropic component cannot be assessed until a detailed thermodynamic study is performed for this complex. This hypothesis of K50 side-chain mobility will be explored further in the future, but for now, it is complementary to the data for the EnQ50K mutant (*7*). The crystal structure indicates that there are two alternate conformations for the K50 side chain, one in which the side chain points to base pairs 5 and 6, and one in which the side chain is oriented slightly more toward base pair 5. It must be pointed out that this X-ray structure was solved at cryogenic temperatures, so there is the possibility that there is a freezing out of a subset of conformational populations. It is possible that these results indicate two static, nearly isoenergetic conformations for this side chain of EnQ50K, rather than a dynamic fluctuation between two conformations. The *B*-factors in this case provide no evidence for distinguishing between these possibilities. The *B*-factors are low for the side chain of K50 in the 1.9 Å crystal structure of EnQ50K, varying over the range 20.8 to 23.6, which are the lowest values in the protein, aside from the aromatic ring of F49. *B*-factors of about 20 indicate uncertainties of about 0.5 Å. Typically, *B*-factors of 60 or greater in high-resolution crystal structures indicate possible mobility of a side chain. So, according to the crystal results, the side-chain position of K50 is well defined in the crystal, in contrast to the possible mobility of the K50 side chain seen in our results. The true nature of the side-chain conformation and dynamics may involve a combination of the states revealed by the two different experimental approaches, so that the K50 side chain has two predominant conformations, and fluctuates between these alternatives.

Although a more detailed characterization of the side-chain dynamics in the PITX2−DNA interface must await data from experimental NMR relaxation measurements and molecular dynamics simulations, substantial support for our observation of flexibility in the K50 side chain already exists from studies of related systems. Significant broadening of side-chain resonances at the protein−DNA interface was observed in studies of homeodomain−DNA complexes of Antennapedia (*97*) and NK-2 (*32*). Moreover, flexibility in lysine side chains appears to be a significant feature of various modes of protein−DNA interactions. Foster and co-workers (*98*) have reported clear indications of substantial, conformational fluctuations in lysine side chains in the interface of the zinc-finger protein TFIIIA with its DNA binding site, including the observation of broadened resonances and multiple NOE contacts that strongly suggest rapid conformational averaging. Significant line-broadening effects were also reported for a lysine side chain in NMR studies of the telomeric DNA complex of trf1 (*99*). In addition to NMR studies, molecular dynamics simulations of wild-type (*31*) and a Q50K mutant (*33*) of the Antennapedia homeodomain bound to DNA provide further evidence in support of a dynamic homeodomain−DNA interface. For example, the Q50K simulations indicated that the side chain of K50 exhibited very pronounced mobility, with several arrangements of the lysine side-chain torsion angles allowing for frequent contacts, both hydrogen-bonding and hydrophobic interactions, with base pairs 5 and 6 in the TAATCC binding site. In this case, the lysine in the Q50K mutant provides both entropic and enthalpic contributions to protein−DNA affinity. A general observation arising from the known structures of homeodomain−DNA complexes is that the region of position 50 is not in intimate contact with the bases of the major groove. Such a relatively unrestrained arrangement allows for relatively long-range contacts to be formed in multiple, possibly isoenergetic ways.

Previous studies have shown that the lysine at position 50 is critical for its binding to the TAATCC DNA binding site (*82*). In contrast, homeodomains with a glutamine at position 50 bind to TAATGG sites with a higher affinity. The glutamine at position 50 appears to have a more modest role. When this residue is mutated to an alanine, the Q50A mutant has an affinity and specificity very similar to those of the wild-type protein, but when mutated to a lysine, the specificity changes (*5*). These studies, along with the current results, indicate that the interaction between K50 and the two guanines at positions 5 and 6 is vital to the affinity and specificity of the protein. The current model for specific homeodomain−DNA interactions consists of a fluctuating network of hydrogen bonds formed between polar groups of the protein and the DNA, and the interfacial water (*31*). These interactions are further complemented by hydrophobic contacts. The possible fluctuating hydrogen-bonding interactions between K50 and the DNA and subsequent strict specificity of this class of homeodomains are consistent with this model. Investigation of side-chain−base interactions has shown that lysine−guanine interactions are very common (*101*). K50 homeodomains may have such a strong specificity for the TAATCC site because the orientation of the lysine is in an ideal position for the charged group to make hydrogen-bonding contacts with the two guanines. In contrast, these hydrogen bonds cannot be made with cytosines, which are in these positions for the Q50 binding site TAATGG (*101*). The N7 of guanine is the most electronegative region of the major groove (*102*), and the favorable interactions that the lysine can make with both guanines in a mobile model may determine why K50 homeodomains are so specific for the TAATCC binding site, rather than other binding sites.

Analysis of Residues Mutated in Rieger Syndrome. There have been 9 mutations found in the PITX2 homeodomain

Table 2. PITX2 Homeodomain Mutations (*35*, *45*, *48*, *103*−*109*)

mutation	disease	properties
L16Q	Rieger syndrome	unstable, no activation, no DNA binding
T30P	Rieger syndrome	no activation, only binds consensus
R31H	iridogoniodysgenesis	reduced activation, only binds consensus site
V45L	Rieger syndrome	<10-fold reduction in DNA binding, 200% increase in activation
R46W	iris hypoplasia	reduced binding to nonconsensus site, reduced activation
K50E	Rieger syndrome	no DNA binding or activation, dominant negative
K50Q	Rieger syndrome	not known
R52C	Rieger syndrome	not known
R53P	Rieger syndrome	no nonconsensus binding, no activation, dominant negative

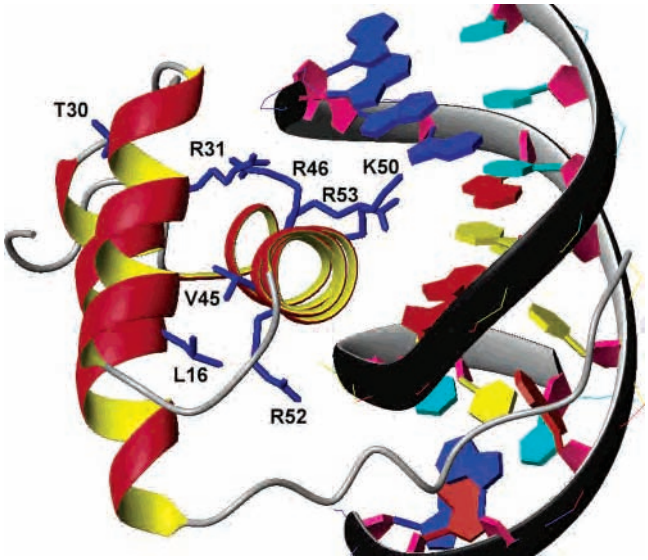

FIGURE 7: Ribbon diagram of the PITX2 homeodomain−DNA complex showing the positions of the side chains for the residues known to be mutated in Rieger syndrome and related disorders.

in Rieger syndrome and related disorders (*35*, *45*, *48*, *103*−*109*). These mutations, along with their known biochemical effects, are listed in Table 2. The consequences of these mutations vary. Some mutations cause a total lack of DNA binding, while others can still bind DNA, albeit with a decreased affinity. These consequences are directly reflected in the severity of the disease. A model of the PITX2 homeodomain structure was created previously by threading analysis, which allowed predictions to be made regarding the role of Rieger syndrome mutations in PITX2 dysfunction, although it is not necessarily an indication of the true molecular structure (*83*).

The orientations of the side chains altered in Rieger syndrome patients are shown in Figure 7. Analysis of these orientations provides insights into the role of each side chain, and how mutations in these positions could alter the structure and function of the protein. Future studies will focus on analyzing the mutant proteins by NMR spectroscopy. The side chain of highly conserved L16 points toward the interior hydrophobic core of the protein, and is probably involved in stabilizing both the formation of this core and the overall tertiary structure of the protein; the L16Q mutation would therefore be expected to destabilize or disrupt this hydrophobic core. The side chain of T30 extends outward from the second helix, away from the DNA, so it does not appear to play a role in DNA recognition. Biochemical studies have shown that this mutant can still bind consensus DNA, but no longer activates transcription of a reporter gene (*25*). This residue may perform an activation function by interacting with other proteins, which could easily be disrupted by the effects of the proline mutation. An interesting observation is that, in many homeodomains, residue 30 is involved in a salt bridge to residue 19, whereas this is not possible for PITX2. The side chain of R31, as described above, appears to contact the DNA backbone phosphate of G82. Therefore mutating this residue, even to another positively charged residue, may disrupt this interaction with the DNA and may disrupt a possible salt bridge with E42. The histidine side chain at this position in the mutant may not have favorable steric interactions with the DNA. The side chain of V45 points toward the interior of the protein from the third helix. Like L16, this side chain appears to be involved in formation of the hydrophobic core of the protein. Unlike the L16Q mutant, the V45L mutant has the unusual characteristic of having a greatly heightened activation function, while having a reduced DNA-binding ability. It is possible that this mutant affects the protein in a way that alters these two functions separately, with a different fold of the protein that allows for a more efficient interaction with other proteins. For example, altered interactions of the PITX2 homeodomain with the C-terminal tail of the full-length PITX2 protein could have differential effects on DNA binding and activation (*110*). The DNA-binding functions of R46, K50, R52, and R53 were discussed in detail above. Mutating these residues would disrupt many favorable interactions with the DNA, and biochemical studies have indicated that these mutations interfere with DNA binding. Overall, these results are similar to the threading analysis, but provide a more direct and detailed understanding of the roles of these residues.

Many of the residues in the PITX2 homeodomain found to be altered in Rieger syndrome are involved in contacting the DNA. Other residues are involved in forming the hydrophobic core of the protein, which stabilizes the global fold. The analysis of mutations causing structural changes could be very relevant for the understanding and prediction of dysfunctions caused by mutations in homeodomains, as several homeodomains are known to be involved in various diseases (*111*−*115*).

CONCLUDING REMARKS

The structure previously determined for the Engrailed Q50K mutant (*7*) provided some interesting insights into the possible role of lysine at position 50. The presence of hydrogen bonds between position 50 and the DNA had not been seen previously. But many questions remained unanswered concerning the role of lysine in a native K50 homeodomain. For example, the Engrailed mutant has a dissociation constant of 0.0088 nM (*7*), representing an unusually high affinity for homeodomain−DNA interactions. Previous studies have indicated that proteins with excessively high affinities for DNA or RNA can cause functional defects (*116*, *117*). The unusually high affinity of EnQ50K for DNA suggests that it may have properties that make it different from natural K50 homeodomains. Unlike the Engrailed mutant, the native K50 class homeodomains PITX2 and Bicoid have properties that make them unstable in free forms,

and have affinities within the normal nanomolar range (*25*, *26*, Supporting Information). When DNA is not present, these proteins will irreversibly aggregate and precipitate out of solution at micromolar concentrations. These differences in biochemical properties between the mutant and natural K50 proteins suggest the importance of understanding the structural properties of lysine at position 50 in the context of a native K50 class protein.

But the question still remains as to what causes these differences. The authors of the EnQ50K structure found that the mutant bound to DNA more tightly and specifically than did the native protein (*7*). They hypothesized that this was due to very specific hydrogen bonds between the K50 side chain and the guanines at positions 5 and 6 on the antisense strand. In our study, we found that the native K50 homeodomain PITX2 has a slightly different tertiary structure, with helix 1 being closer to helix 2 than in other homeodomains, including the EnQ50K mutant. Helix 3 is angled about 0.5 Å closer to the N-terminus of helix 1 and C-terminus of helix 2 than EnQ50K. This appears to cause a difference in the way that helix 1 and helix 3 can interact, and previous studies have shown that this interaction between the helices stabilizes the global fold of the homeodomain (*3*, *81*, *82*). Another Q50K mutant, this time of Fushi tarazu, is unable to bind nonconsensus DNA sites that PITX2 and Bicoid are able to recognize (*21*). It is currently unknown whether the Engrailed mutant can bind nonconsensus sites. These differences in affinity and specificity may involve any of the differing residues between these homeodomains. Positions 50 and 54 have been shown to be involved in recognizing nonconsensus DNA sites (*22*), and it is possible that other residues are also involved. Within the third helix, position 52 of Engrailed is a lysine. In PITX2, Bicoid, and Fushi tarazu, this residue is an arginine. We do not know whether having lysine residues at both positions 50 and 52 could contribute to the unnaturally tight binding of EnQ50K, but this is a possibility.

The current study of the solution structure of the PITX2 homeodomain reveals possible fluctuating interactions between the K50 side chain and the DNA. It is possible that this mobile side chain may allow the protein to sample multiple DNA binding sites, and enable binding to the nonconsensus sites, though at a slightly lower affinity. It will be interesting in the future to determine if other natural K50 class proteins share similar properties with PITX2. Future studies will focus on analyzing Rieger mutants of the PITX2 homeodomain, and analyzing the structural features of this protein when bound to nonconsensus DNA binding sites. This will allow a greater understanding of the roles of specific residues in consensus and nonconsensus DNA binding, and a greater understanding of how proteins can recognize multiple DNA sites to activate transcription of genes.

ACKNOWLEDGMENT

We would like to especially thank Dr. Jeffery C. Murray and Dr. Elena V. Semina from the University of Iowa for providing the original *pitx2* clone. We would also like to thank Dr. Jack Howarth for computer support and maintaining the NMR facility at the University of Cincinnati.

NOTE ADDED AFTER ASAP PUBLICATION

This paper was originally published 4/26/05 with an incorrect entry in Table 1. The corrected version was also published 4/26/05.

SUPPORTING INFORMATION AVAILABLE

Sequence alignment of homeodomains (Table S1), binding affinity of the PITX2 homeodomain (Figure S1), HSQC with the residues labeled (Figure S2), chemical shift assignments for the PITX2 homeodomain and its DNA binding site (Tables S2-S7), protein−DNA contacts (Table S8), and Ramachandran plot for the PITX2 homeodomain (Figure S3). This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Gehring, W. J., Muller, M., Affolter, M., Percival-Smith, A., Billeter, M., Qian, Y. Q., Otting, G., and Wüthrich, K. (1990) The structure of the homeodomain and its functional implications, *Trends Genet. 6*, 323−329.
2. Dave, V., Zhao, C., Yang, F., Tung, C., and Ma, J. (2000) Reprogrammable recognition codes in Bicoid homeodomain-DNA interaction, *Mol. Cell. Biol. 20*, 7673−7684.
3. Gehring, W. J., Affolter, M., and Burglin, T. (1994) Homeodomain Proteins, *Annu. Rev. Biochem. 63*, 487−526.
4. Treisman, J., Harris, E., Wilson, D., and Desplan, C. (1992) The Homeodomain: A New Face for the Helix-Turn-Helix?, *BioEssays 14*, 145−150.
5. Wolberger, C. (1993) Transcription factor structure and DNA binding, *Curr. Opin. Struct. Biol. 3*, 3−10.
6. Wilson, D. S., Sheng, G., Jun, S., and Desplan, C. (1996) Conservation and diversification in homeodomain-DNA interactions: A comparative genetic analysis, *Proc. Natl. Acad. Sci. U.S.A. 93*, 6886−6891.
7. Tucker-Kellogg, L., Rould, M. A., Chambers, K. A., Ades, S. E., Sauer, R. T., and Pabo, C. O. (1997) Engrailed (Gln50→Lys) homeodomain-DNA complex at 1.9 Å resolution: structural basis for enhanced affinity and altered specificity, *Structure 5*, 1047−1054.
8. Hanes, S. D., and Brent, R. (1989) DNA Specificity of the Bicoid activator protein is determined by homeodomain recognition helix residue 9, *Cell 57*, 1275−1283.
9. Ades, S. E., and Sauer, R. T. (1994) Differential DNA-binding specificity of the Engrailed homeodomain: The role of residue 50, *Biochemistry 33*, 9187−9194.
10. Fraenkel, E., Rould, M. A., Chambers, K. A., and Pabo, C. O. (1998) Engrailed homeodomain-DNA complex at 2.2 Å resolution: A detailed view of the interface and comparison with other engrailed structures, *J. Mol. Biol. 284*, 351−361.
11. Grant, R. A., Rould, M. A., Klemm, J. D., and Pabo, C. O. (2000) Exploring the Role of Glutamine 50 in the Homeodomain-DNA Interface: Crystal Structure of Engrailed (Gln50→Ala) Complex at 2.0 Å, *Biochemistry 39*, 8187−8192.
12. Percival-Smith, A., Muller, M., Affolter, M., and Gehring, W. J. (1990) The Interaction with DNA of Wild-Type and Mutant *Fushi tarazu* Homeodomains, *EMBO J. 9*, 3967−3974.
13. Tron, A. E., Bertoncini, C. W., Palena, C. M., Chan, R. L., and Gonzalez, D. H. (2001) Combinatorial Interactions of Two Amino Acids with a Single Base Pair Define Target Site Specificity in Plant Dimeric Homeodomain Proteins, *Nucleic Acids Res. 29*, 4866−4872.
14. Pomerantz, J. L., and Sharp, P. A. (1994) Homeodomain determinants of major groove recognition, *Biochemistry 33*, 10851−10858.
15. Banerjee-Basu, S., Moreland, T., Hsu, B. J., Trout, K. L., and Baxevanis, A. D. (2003) The homeodomain resource: 2003 update, *Nucleic Acids Res. 31*, 304−306.
16. Otting, G., Qian, Y. Q., Muller, M., Affolter, M., Gehring, W., and Wüthrich, K. (1988) Secondary structure determination for the Antennapedia homeodomain by nuclear magnetic resonance and evidence for a helix-turn-helix motif, *EMBO 7*, 4305−4309.

17. Li, T., Stark, M. R., Johnson, A. D., and Wolberger, C. (1995) Crystal Structure of the MATa1/MATα2 Homeodomain Heterodimer Bound to DNA, *Science 270*, 262−269.
18. Cox, M., van Tilborg, P. J., de Laat, W., Boelens, R., van Leeuwen, H. C., van der Vliet, P. C., and Kaptein, R. (1995) Solution structure of the Oct-1 POU homeodomain determined by NMR and restrained molecular dynamics, *J. Biomol. NMR 6*, 23−32.
19. Piper, D. E., Batchelor, A. H., Chang, C., Cleary, M. L., and Wolberger, C. (1999) Structure of a HoxB1-Pbx1 Heterodimer Bound to DNA: Role of the Hexapeptide and a Fourth Homeodomain Helix in Complex Formation, *Cell 96*, 587−597.
20. Tejada, M. L., Jia, Z., May, D., and Deeley, R. G. (1999) Determinants of the DNA-binding specificity of the Avian homeodomain protein, AKR, *DNA Cell Biol. 18*, 791−804.
21. Zhao, C., Dave, V., Yang, F., Scarborough, T., and Ma, J. (2000) Target Selectivity of Bicoid is Dependent on Nonconsensus Site Recognition and Protein-Protein Interaction, *Mol. Cell. Biol. 20*, 8112−8123.
22. Pellizzari, L., Tell, G., Fabbro, D., Pucillo, C., and Damante, G. (1997) Functional interference between contacting amino acids of homeodomains, *FEBS Lett. 407*, 320−324.
23. Clarke, N. D. (1995) Covariation of residues in the homeodomain sequence family, *Protein Sci. 4*, 2269−2278.
24. Simon, M. D., Sato, K., Weiss, G. A., and Shokat, K. M. (2004) A phage display selection of engrailed homeodomain mutants and the importance of residue Q50, *Nucleic Acids Res. 32*, 3623−3631.
25. Amendt, B. A., Sutherland, L. B., Semina, E. V., and Russo, A. F. (1998) The Molecular Basis of Rieger Syndrome, *J. Biol. Chem. 273*, 20066−20072.
26. Ma, X., Yuan, D., Diepold, K., Scarborough, T., and Ma, J. (1996) The *Drosophila* morphogenetic protein Bicoid binds DNA cooperatively, *Development 122*, 1195−1206.
27. Yuan, D., Ma, X., and Ma, J. (1999) Recognition of Multiple Patterns of DNA Sites by Drosophila Homeodomain Protein Bicoid, *J. Biochem. 125*, 809−817.
28. Hjalt, T. A., Amendt, B. A., and Murray, J. C. (2001) PITX2 Regulates Procollagen Lysyl Hydroxylase (PLOD) Gene Expression: Implications for the Pathology of Rieger Syndrome, *J. Cell Biol. 152*, 545−552.
29. Espinoza, H. M., Cox, C. J., Semina, E. V., and Amendt, B. A. (2002) A molecular basis for differential developmental anomalies in Axenfeld-Rieger syndrome, *Hum. Mol. Genet. 11*, 743−753.
30. Hirsch, J. A., and Aggarwal, A. K. (1995) Structure of the even-skipped homeodomain complexed to AT-rich DNA: new perspectives on homeodomain specificity, *EMBO J. 14*, 6280−6291.
31. Billeter, M., Guntert, P., Luginbuhl, P., and Wüthrich, K. (1996) Hydration and DNA Recognition by Homeodomains, *Cell 85*, 1057−1065.
32. Tsao, D. H., Gruschus, J. M., Wang, L. H., Nirenburg, M., and Ferretti, J. A. (1994) Elongation of helix III of the NK-2 homeodomain upon binding to DNA: a secondary structure study by NMR, *Biochemistry 33*, 15053−15060.
33. Gutmanas, A., and Billeter, M. (2004) Specific DNA recognition by the *Antp* homeodomain: MD simulations of specific and nonspecific complexes, *Proteins: Struct. Funct., Bioinf. 57*, 772−782.
34. Duan, J., and Nilsson, L. (2002) The role of residue 50 and hydration water molecules in homeodomain DNA recognition, *Eur. Biophys. J. 31*, 306−316.
35. Semina, E. V., Reiter, R., Leysenns, N. J., Alward, W. L., Small, K. W., Datson, N. A., Siegel-Bartelt, J., Bierke-Nelson, D., Bitoun, P., Zabel, B. U., Carey, J. C., and Murray, J. C. (1996) Cloning and characterization of a novel bicoid-related homeobox transcription factor gene, RIEG, involved in Rieger syndrome, *Nat. Genet. 14*, 392−398.
36. Gage, P. J., and Camper, S. A. (1997) Pituitary homeobox 2, a novel member of the bicoid-related family of homeobox genes, is a potential regulator of anterior structure formation, *Hum. Mol. Genet. 6*, 457−464.
37. Mucchielli, M., Mitsiadis, T. A., Raffo, S., Brunet, J. F., Proust, J. P., and Goridis, C. (1997) Mouse Otlx2/RIEG expression in the odontogenic epithelium precedes tooth initiation and requires mesenchyme-derived signals for its maintenance, *Dev. Biol. 189*, 275−284.
38. Hjalt, T. A., Semina, E. V., Amendt, B. A., and Murray, J. C. (2000) The PITX2 protein in mouse development, *Dev. Dyn. 218*, 195−200.
39. Arakawa, H., Nakamura, T., Zhadanov, A. B., Fidanza, V., Yano, T., Bullrich, F., Shimizu, M., Blechman, J., Mazo, A., Canaani, E., and Croce, C. M. (1998) Identification and Characterization of the ARP1 Gene, a Target for the human acute leukemia ALL1 gene, *Proc. Natl. Acad. Sci. U.S.A. 95*, 4573−4578.
40. Kitamura, K., Miura, H., Yanazawa, M., Miyashita, T., and Kato, K. (1997) Expression patterns of Brx1 (Rieg gene), Sonic hedgehog, Nkx2.2, Dlx1 and Arx during zona limitans intrathalamica and embryonic ventral lateral geniculate nuclear formation, *Mech. Dev. 67*, 83−96.
41. Gage, P. J., Suh, H., and Camper, S. A. (1999) Dosage requirement of PITX2 for development of multiple organs, *Development 126*, 4643−4651.
42. Lu, M.-F., Pressman, C., Dyer, R., Johnson, R. L., and Martin, J. F. (1999) Function of Rieger syndrome gene in left-right asymmetry and craniofacial development, *Nature 401*, 276−278.
43. Lin, C. R., Kioussi, C., O'Connell, S., Briata, P., Szeto, D., Liu, F., Izpisua-Belmonte, J. C., and Rosenfeld, M. G. (1999) PITX2 regulates lung asymmetry, cardiac positioning and pituitary and tooth morphogenesis, *Nature 401*, 279−282.
44. Xia, K., Wu, L., Liu, X., Xi, X., Liang, D., Zheng, D., Cai, F., Pan, Q., Long, Z., Dai, H., Hu, Z., Tang, B., Zhang, Z., and Xia, J. (2004) Mutation in PITX2 is associcated with ring dermoid cornea, *J. Med. Genet. 41*, e129.
45. Priston, M., Kozlowski, K., Gill, D., Letwin, K., Buys, Y., Levin, A. V., Walter, M. A., and Heon, E. (2001) Functional analyses of two newly identified PITX2 mutants reveal a novel molecular mechanism for Axenfeld-Rieger syndrome, *Hum. Mol. Genet. 10*, 1631−1638.
46. Lines, M. A., Kozlowski, K., Kulak, S. C., Allingham, R. R., Heon, E., Ritch, R., Levin, A. V., Shields, M. B., Damji, K. F., Newlin, A., and Walter, M. A. (2004) Characterization and prevalence of PITX2 microdeletions and mutations in Axenfeld-Rieger malformations, *Invest. Ophthalmol. Vis. Sci. 45*, 828−833.
47. Phillips, J. C. (2002) Four novel mutations in the PITX2 gene in patients with Axenfeld-Rieger syndrome, *Ophthalmic Res. 34*, 324−326.
48. Kulak, S. C., Kozlowski, K., Semina, E. V., Pearce, W. G., and Walter, M. A. (1998) Mutation in the RIEG1 gene in patients with iridogoniodysgenesis syndrome, *Hum. Mol. Genet. 7*, 1113−1117.
49. Grzesiek, S., and Bax, A. (1992) Improved 3D triple-resonance NMR techniques applied to a 31-Kda Protein, *J. Magn. Reson. 96*, 432−440.
50. Ikura, M., Kay, L. E., and Bax, A. (1990) A novel-approach for sequential assignment of H-1, C-13, and N-15 spectra of larger proteins: heteronuclear triple-resonance 3-dimensional NMR spectroscopy: application to calmodulin, *Biochemistry 29*, 4659−4667.
51. Kay, L. E., Xu, G. Y., and Yamazaki, T. (1994) Enhanced-sensitivity triple-resonance spectroscopy with minimal H_2O saturation, *J. Magn. Reson., Ser. A 109*, 129−133.
52. Muhandiram, D. R., and Kay, L. E. (1994) Gradient-enhanced triple-resonance 3-dimensional NMR experiments with improved sensitivity, *J. Magn. Reson., Ser. B 103*, 203−216.
53. Grzesiek, S., and Bax, A. (1992) Correlating backbone amide and side-chain resonances in larger proteins by multiple relayed triple resonance NMR, *J. Am. Chem. Soc. 114*, 6291−6293.
54. Wittekind, M., and Mueller, L. (1993) HNCACB, a high-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the α-carbon and β-carbon resonances in proteins, *J. Magn. Reson., Ser. B 101*, 201−205.
55. Sattler, M., Schleucher, J., and Griesinger, C. (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients, *Prog. NMR Spectrosc. 34*, 93−158.
56. Zhang, O., Kay, L. E., Olivier, J. P., and Forman-Kay, J. D. (1994) Backbone ^1H and ^{15}N resonance assignments of the N-terminal SH3 domain of *drk* in folded and unfolded states using enhanced-sensitivity pulsed field gradient NMR techniques, *J. Biomol. NMR 4*, 845−858.
57. Kay, L. E., Keifer, P., and Saarinen, T. (1992) Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity, *J. Am. Chem. Soc. 114*, 10663−10665.
58. Marion, D., Driscoll, P. C., Kay, L. E., Wingfield, P. T., Bax, A., Gronenborn, A. M., and Clore, G. M. (1989) Overcoming the overlap problem in the assignment of H-1-NMR spectra of larger proteins by use of 3-dimensional heteronuclear H-1-N-15 Hart-

mann-Hahn multiple quantum coherence and nuclear Overhauser multiple quantum coherence spectroscopy-application to interleukin-1-beta, *Biochemistry 28*, 6150−6156.
59. Zerbe, O., Szyperski, T., Ottiger, M., and Wüthrich, K. (1996) Three-dimensional ^{1}H-TOCSY-relayed ct-[^{13}C,^{1}H]-HMQC for aromatic spin system identification in uniformly ^{13}C-labeled proteins, *J. Biomol. NMR 7*, 99−106.
60. Vuister, G. W., and Bax, A. (1993) Quantitative J correlations: A new approach for measuring homonuclear three-bond J($H^{N}H^{\alpha}$) coupling constants in ^{15}N-enriched proteins, *J. Am. Chem. Soc. 115*, 7772−7777.
61. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley & Sons, New York.
62. Otting, G., and Wüthrich, K. (1990) Heteronuclear filters in two-dimensional [^{1}H,^{1}H]-NMR spectroscopy: combined use with isotope labeling for studies of macromolecular conformation and intermolecular interactions, *Q. Rev. Biophys. 23*, 39−96.
63. Breeze, A. L. (2000) Isotope-filtered NMR methods for the study of biomolecular structure and interactions, *Prog. NMR Spectrosc. 36*, 323−372.
64. Talluri, S., and Wagner, G. (1996) An optimized 3D NOESY-HSQC, *J. Magn. Reson., Ser. B 112*, 200−205.
65. Stuart, A. C., Borzilleri, K. A., Withka, J. M., and Palmer, A. G. (1999) Compensating for variations in ^{1}H−^{13}C scalar coupling constants in isotope-filtered NMR experiments, *J. Am. Chem. Soc. 121*, 5346−5347.
66. Lee, W., Revington, M. J., Arrowsmith, C., and Kay, L. E. (1994) A pulsed field gradient isotope-filtered 3D ^{13}C HMQC-NOESY experiment for extracting intermolecular NOE contacts in molecular complexes, *FEBS Lett. 350*, 87−90.
67. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes, *J. Biomol. NMR 6*, 277−293.
68. Goddard, T. D., Kneller, D. G. *SPARKY 3*, University of California, San Francisco.
69. Güntert, P., Mumenthaler, C., and Wüthrich, K. (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA, *J. Mol. Biol. 273*, 283−298.
70. Herrmann, T., Güntert, P., and Wüthrich, K. (2002) Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA, *J. Mol. Biol. 319*, 209−227.
71. Case, D. A., Pearlman, D. A., Caldwell, J. W., Cheatham, T. E., Ross, W. S., Simmerling, C. L., Darden, T. A., Merz, K. M., Stanton, R. V., Cheng, A. L., Vincent, J. J., Crowley, M., Tsui, V., Radmer, R. J., Duan, Y., Pitera, J., Massova, I., Seibel, G. L., Singh, U. C., Weiner, P. K., and Kalman, P. A. (1996) *AMBER7*, University of California, San Francisco.
72. Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. S., Cheatham, T. E., deBolt, S., Ferguson, D., Seibel, G., and Kollman, P. A. (1995) AMBER, a computer program for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to elucidate the structures and energies of molecules, *Comput. Phys. Commun. 91*, 1−41.
73. Wuttke, D. S., Foster, M. P., Case, D. A., Gottesfeld, J. M., and Wright, P. E. (1997) Solution structure of the first three zinc fingers of TFIIIA bound to the cognate DNA sequence: Determinants of affinity and sequence specificity, *J. Mol. Biol. 273*, 183−206.
74. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures, *J. Appl. Crystallogr. 26*, 283−291.
75. Laskowski, R. A., Rullmann, J. A. C., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR, *J. Biomol. NMR 8*, 477−486.
76. Konradi, R., Billeter, M., and Wüthrich, K. (1996) MOLMOL: a program for display and analysis of macromolecular structures, *J. Mol. Graphics 14*, 51−55.
77. Spera, S., and Bax, A. (1991) Empirical correlation between protein backbone conformation and C^{α} and C^{β} ^{13}C nuclear magnetic resonance shifts, *J. Am. Chem. Soc. 113*, 5490−5492.
78. Luginbuhl, P., Szyperski, T., and Wüthrich, K. (1995) Statistical basis for the use of $^{13}C^{\alpha}$ chemical shifts in protein structure determination, *J. Magn. Reson. B 109*, 229−233.
79. Scott, M. P., Tamkun, J. W., and Hartzell, G. W. (1989) The structure and function of the homeodomain, *Biochim. Biophs. Acta 989*, 25−48.
80. Billeter, M. (1996) Homeodomain-type DNA recognition, *Prog. Biophys. Mol. Biol. 66*, 211−225.
81. Qian, Y. Q., Billeter, M., Otting, G., Muller, M., Gehring, W. J., and Wüthrich, K. (1989) The Structure of the Antennapedia Homeodomain Determined by NMR Spectroscopy in Solution: Comparison with Prokaryotic Repressors, *Cell 59*, 573−580.
82. Kornberg, T. B. (1993) Understanding the Homeodomain, *J. Biol. Chem. 268*, 26813−26816.
83. Banerjee-Basu, S., and Baxevanis, A. (1999) Threading analysis of the PITX2 homeodomain: predicted structural effects of mutations causing Rieger Syndrome and Iridogoniodysgenesis, *Hum. Mutat. 14*, 312−319.
84. Billeter, M., Qian, Y. Q., Otting, G., Muller, M., Gehring, W. J., and Wüthrich, K. (1993) Determination of the nuclear magnetic resonance solution structure of an Antennapedia homeodomain-DNA complex, *J. Mol. Biol. 234*, 1084−1093.
85. Kissinger, C. R., Liu, B. S., Martin-Blanco, E., Kornberg, T. B., and Pabo, C. O. (1990) Crystal structure of an engrailed homeodomain-DNA complex at 2.8 Å resolution: a framework for understanding homeodomain-DNA interactions, *Cell 63*, 579−590.
86. Qian, Y. Q., Furukubo-Tokunaga, K., Resendez-Perez, D., Muller, M., Gehring, W. J., and Wüthrich, K. (1994) Nuclear magnetic resonance solution structure of the *Fushi tarazu* homeodomain from *Drosophila* and comparison with the *Antennapedia* homeodomain, *J. Mol Biol. 238*, 333−345.
87. Gruschus, J. M., Tsao, D. H., Wang, L. H., Nirenberg, M., and Ferretti, J. A. (1997) Interactions of the vnd/NK-2 homeodomain with DNA by nuclear magnetic resonance spectroscopy: basis of binding specificity, *Biochemistry 36*, 5372−5380.
88. Wolberger, C., Vershon, A. K., Liu, B., Johnson, A. D., and Pabo, C. O. (1991) Crystal structure of a MAT alpha 2 homeodomain-operator complex suggests a general model for homeodomain-DNA interactions, *Cell 67*, 517−528.
89. Damante, G., Tell, G., Leonardi, A., Fogolari, F., Bortolotti, N., DiLauro, R., and Formisano, S. (1994) Analysis of the conformation and stability of rat TTF-1 homeodomain by circular dichroism, *FEBS. Lett. 354*, 293−296.
90. Carra, J. H., and Privalov, P. L. (1997) Energetics of folding and DNA binding of the MAT alpha 2 homeodomain, *Biochemistry 36*, 526−535.
91. Otting, G., Qian, Y. Q., Muller, M., Affolter, M., Gehring, W., and Wüthrich, K. (1988) Secondary structure determination for the *Antennapedia* homeodomain by nuclear magnetic resonance and evidence for a helix-turn-helix motif, *EMBO J. 7*, 4305−4309.
92. Yamamoto, K., Yee, C. C., Shirakawa, M., and Kyogoku, Y. (1992) Characterization of the bacterially expressed Drosophila engrailed homeodomain, *J. Biochem. Tokyo 111*, 793−797.
93. Iurcu-Mustata, G., Van Belle, D., Wintjens, R., Prévost, M., and Rooman, M. (2001) Role of salt bridges in homeodomains investigated by structural analyses and molecular dynamics simulations, *Biopolymers 59*, 145−159.
94. Jones, S., van Heyningen, P., Berman, H. M., and Thornton, J. M. (1999) Protein-DNA interactions: A structural analysis, *J. Mol. Biol. 287*, 877−896.
95. Fernandez, C., Szyperski, T., Billeter, M., Ono, A., Iwai, H., Kainosho, M., and Wüthrich, K. (1999) Conformational changes of the BS2 operator DNA upon complex formation with the *Antennapedia* homeodomain studied by NMR with 13C/15N-labeled DNA, *J. Mol. Biol. 292*, 609−617.
96. McDonald, I. K., and Thornton, J. M. (1994) Satisfying Hydrogen Bonding Potential in Proteins, *J. Mol. Biol. 238*, 777−793.
97. Qian, Y. Q., Otting, G., Billeter, M., Müller, M., Gehring, W., and Wüthrich, K. (1993) Nuclear magnetic resonance spectroscopy of a DNA complex with the uniformly 13C-labeled *Antennapedia* homeodomain and structure determination of the DNA-bound homeodomain, *J. Mol. Biol. 234*, 1070−1083.
98. Foster, M. P., Wuttke, D. S., Radhakrishnan, I., Case, D. A., Gottesfeld, J. M., and Wright, P. E. (1997) Domain packing and dynamics in the DNA complex of the N-terminal zinc fingers of TFIIIA, *Nat. Struct. Biol. 4*, 605−608.
99. Nishikawa, T., Okamura, H., Nagadoi, A., König, P., Rhodes, D., and Nishimura, Y. (2001) Solution structure of a telomeric DNA complex of human TRF1, *Structure 9*, 1237−1251.
100. Doig, A. J., and Sternberg, M. J. E. (1995) Side-chain conformational entropy in protein folding, *Protein Sci. 4*, 2247−2251.
101. Mandel-Gutfreund, Y., Schueler, O., and Margalit, H. (1995) Comprehensive analysis of hydrogen bonds in regulatory protein-

DNA complexes: in search of common principles, *J. Mol. Biol. 253*, 370−382.
102. Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer-Verlag, New York.
103. Saadi, I., Semina, E. V., Amendt, B. A., Harris, D. J., Murphy, K. P., Murray, J. C., and Russo, A. F. (2001) Identification of a Dominant Negative Homeodomain Mutation in Rieger Syndrome, *J. Biol. Chem. 276*, 23034−23041.
104. Heon, E., Sheth, B. P., Kalenak, J. W., Sunden, S. L., Streb, L. M., Taylor, C. M., Alward, W. L., Sheffield, V. C., and Stone, E. M. (1995) Linkage of autosomal dominant iris hypoplasia to the region of the Rieger Syndrome locus (4q25), *Hum. Mol. Genet. 4*, 1435−1439.
105. Alward, W. L. M., Semina, E. V., Kalenak, J. W., Heon, E., Sheth, B. P., Stone, E. M., and Murray, J. C. (1998) Autosomal Dominant Iris Hypoplasia is Caused by a Mutation in the Rieger Syndrome (RIEG/PITX2) Gene, *Am. J. Opthalmol. 125*, 98−100.
106. Chisholm, E. A., and Chudley, A. E. (1983) Autosomal dominant iridogoniodysgenesis with associated somatic anomalies: four-generation family with Rieger's syndrome, *Br. J. Ophthalmol. 67*, 529−534.
107. Walter, M. A., Mirzayans, F., Mears, A. J., Hickey, K., and Pearce, W. G. (1996) Autosomal-dominant iridogoniodysgenesis and Axenfeld-Rieger syndrome are genetically distinct, *Ophthalmology 103*, 1907−1915.
108. Murray, J. C., Bennett, S. R., Kwitek, A. E., Small, K. W., Schinzel, A., Alward, W. L., Weber, J. L., Bell, G. I., and Buetow, K. H. (1992) Linkage of Rieger Syndrome to the region of the epidermal growth factor gene on chromosome 4, *Nat. Genet. 2*, 46−49.
109. Quentien, M., Pitoia, F., Gunz, G., Guillet, M., Enjalbert, A., and Pellegrini, I. (2002) Regulation of Prolactin, GH, and Pit-1 Gene Expression in Anterior Pituitary by PITX2: An Approach Using PITX2 Mutants, *Endocrinology 143*, 2839−2851.
110. Amendt, B. A., Sutherland, L. B., and Russo, A. F. (1999) Multifunctional role of the Pitx2 homeodomain protein C-terminal tail, *Mol. Cell. Biol. 19*, 7001−7010.
111. Boncinelli, E. (1997) Homeobox genes and disease, *Curr. Opin. Genet. Dev. 7*, 331−337.
112. Muragaki, Y., Mundlos, S., Upton, J., and Olsen, B. R. (1996) Altered growth and branching patterns in synpolydactyly caused by mutations in HOXD13, *Science 272*, 548−551.
113. Nakamura, T., Largaespada, D. A., Lee, M. P., Johnson, L. A., Ohyashiki, K., Toyama, K., Chen, S. J., Willman, C. L., Chen, I. M., Feinberg, A. P., Jenkins, N. A., Copeland, N. G., and Shaughnessy, J. D., Jr. (1996) Fusion of the nucleoporin gene NUP98 to HOXA9 by the chromosome translocation t(7;11)-(p15;p15) in human myeloid leukaemia, *Nat. Genet. 12*, 154−158.
114. D'Elia, A. V., Tell, G., Paron, I., Pellizzari, L., Lonigro, R., and Damante, G. (2001) Missense Mutations of Human Homeoboxes: A Review, *Hum. Mutat. 18*, 361−374.
115. Borrow, J., Shearman, A. M., Stanton, V. P., Jr., Becher, R., Collins, T., Williams, A. J., Dube, I., Katz, F., Kwong, Y. L., Morris, C., Ohyashiki, K., Toyama, K., Rowley, J., and Housman, D. E. (1996) The t(7;11)(p15;p15) translocation in acute myeloid leukaemia fuses the genes for nucleoporin NUP98 and class I homeoprotein HOXA9, *Nat. Genet. 12*, 159−167.
116. Watanabe, K., and Lambowitz, A. M. (2004) High-affinity binding site for a group II intron-encoded reverse transcriptase/maturase within a stem-loop structure in the intron RNA, *RNA 10*, 1433−1443.
117. Monsalve, M., Calles, B., Mencia, M., Rojo, F., and Salas, M. (1998) Binding of phage Φ29 Protein p4 to the Early A2c Promoter: Recruitment of a Repressor by the RNA Polymerase, *J. Mol. Biol. 283*, 559−569.

BI0473253

doi:10.1016/j.jmb.2005.12.007 J. Mol. Biol. (2006) 356, 1137–1151

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

The Solution Structure of the Native K50 Bicoid Homeodomain Bound to the Consensus TAATCC DNA-binding Site

Jamie M. Baird-Titus[1], Kimber Clark-Baldwin[1], Vrushank Dave[2] Carol A. Caperelli[3], Jun Ma[2] and Mark Rance[1]*

[1]*Department of Molecular Genetics, Biochemistry, and Microbiology, University of Cincinnati College of Medicine 231 Albert Sabin Way, Medical Sciences Building, Cincinnati OH 45267-0524, USA*

[2]*Division of Developmental Biology, Children's Hospital Research Foundation, 3333 Burnet Avenue, Cincinnati OH 45229, USA*

[3]*Division of Pharmaceutical Sciences, College of Pharmacy University of Cincinnati Medical Center, Cincinnati OH 45267, USA*

**Corresponding author*

The solution structure of the homeodomain of the *Drosophila* morphogenic protein Bicoid (Bcd) complexed with a TAATCC DNA site is described. Bicoid is the only known protein that uses a homeodomain to regulate translation, as well as transcription, by binding to both RNA and DNA during early *Drosophila* development; in addition, the Bcd homeodomain can recognize an array of different DNA sites. The dual functionality and broad recognition capabilities signify that the Bcd homeodomain may possess unique structural/dynamic properties. Bicoid is the founding member of the K50 class of homeodomain proteins, containing a lysine residue at the critical 50th position (K50) of the homeodomain sequence, a residue required for DNA and RNA recognition; Bcd also has an arginine residue at the 54th position (R54), which is essential for RNA recognition. Bcd is the only known homeodomain with the K50/R54 combination of residues. The Bcd structure indicates that this homeodomain conforms to the conserved topology of the homeodomain motif, but exhibits a significant variation from other homeodomain structures at the end of helix 1. A key result is the observation that the side-chains of the DNA-contacting residues K50, N51 and R54 all show strong signs of flexibility in the protein–DNA interface. This finding is supportive of the adaptive-recognition theory of protein–DNA interactions.

Keywords: bicoid; homeodomain; DNA/RNA-binding protein; NMR; molecular dynamics

Introduction

Homeodomains are an evolutionarily conserved class of DNA-binding domains that are ubiquitous in multicellular organisms.[1–6] Over 1060 unique homeodomain sequences from 112 different species of plants and animals have been isolated and sequenced.[1] They are found in transcriptional regulatory proteins, such as engrailed (fly), MATα (yeast), and PITX2 (human), and are involved in various processes, including the spatial and temporal delineation of developmental regions in embryo,[4] cell-type specification and differentiation,[7] and the maintenance of embryonic stem cells.[8] Bicoid is a homeodomain-containing *Drosophila* transcription factor that directs formation of the anterior–posterior axis in the developing embryo,[9–12] through recognition of enhancer elements of gap and pair-rule genes, such as hunchback, knirps, and even-skipped.[13–18]

The broad goal of understanding the basic mechanisms of DNA information storage/retrieval and, consequently, the basis by which proteins recognize distinct DNA sequences, is evident in the history of homeodomain research. The prevalence of the homeodomain, combined with its small size (60 amino acid residues) and functionality in the absence of the rest of the protein, have placed the homeodomain in a unique position to serve as

Present address: V. Dave, Division of Pulmonary Biology, Children's Hospital Research Foundation, 3333 Burnet Avenue, Cincinnati, OH 45229, USA.

Abbreviations used: NOESY, nuclear Overhauser effect spectroscopy; TOCSY, total correlated spectroscopy; TEV, tobacco etch virus; HSQC, heteronuclear single quantum coherence; HMQC, heteronuclear multiple quantum coherence.

E-mail address of the corresponding author: mark.rance@uc.edu

0022-2836/$ - see front matter

a valuable tool for probing the basis of protein–DNA interactions, and have made it a long-standing subject of functional,[19–33] and structural studies, by both NMR[1,34–46] and X-ray crystallography.[1,47–60] These studies have revealed a conserved global fold consisting of three α helices and a flexible N-terminal arm that becomes more ordered upon DNA binding.[46] Sequence-specific recognition of the DNA-binding site (a double-stranded DNA hexamer defined loosely by a TAAT core in the sense strand)[61] is mediated by amino acid residues located in the third "recognition helix",[55,62] that bind to bases in the major groove, and the N-terminal arm, which wraps around the DNA double helix and makes contacts with bases in the adjacent minor groove.[55,60] However, despite numerous studies showing the conserved global fold and general binding orientation of homeodomains on the DNA, fundamental questions remain regarding the specific nature of interactions between amino acid side-chains and nucleic acids during recognition of binding sites.[36,60]

Previous research has shown that position 50 of the homeodomain, located in the "recognition" helix, plays an important role during recognition of specific DNA-binding sites. This position is occupied most frequently by glutamine but can be occupied by alanine, serine, cysteine, lysine or isoleucine, among others. Position 50 plays a fundamental role in recognizing the bases immediately 3′ to the TAAT core (TAATNN) and, to a lesser extent, position 4 of the TAAT core (TAATNN), and has been the focus of many structural[36,50,52,58] and functional studies.[15,19,21,26,30,63] Homeodomains with glutamine at position 50 (Q50) recognize TAATTA, TAATTG, or TAATGG sites,[29,64] while those with lysine at position 50 (K50), like Bicoid, recognize TAATCC or TAAGCT. Two structural studies of K50 homeodomains have been done, one by Tucker-Kellogg *et al*. of the engrailed Q50K mutant crystal structure,[58] and the second, a recently published solution structure of the native K50 PITX2 homeodomain by our group.[36] Binding studies of the engrailed Q50K mutant indicated that the presence of lysine at position 50 conferred a DNA-binding affinity in the picomolar range ($K_D = 8.8 \times 10^{-12}$ M), considerably higher when compared to other non-K50 homeodomains (K_D range $\sim 10^{-9}$–10^{-10} M).[20,24,27,31,42] Tucker-Kellogg *et al*. associated this increase in binding affinity with the specific hydrogen bond contacts made between K50 and the DNA in their crystal structure.[58] Binding analyses performed with the Bicoid and PITX2 native K50 homeodomains revealed that the two native K50 homeodomains retained binding affinities similar to those seen in other homeodomains (see Chaney *et al*.[36] and see S1 of the Supplementary Data), in contrast to the results reported by Tucker-Kellogg *et al*.,[58] and indicating that the engrailed Q50K homeodomain may not reflect the behavior of native K50 homeodomains.

Additional evidence for the complex role of the position 50 residue comes from the evolutionary co-variation that has been observed among the DNA-contacting amino acid residues of the recognition helix, specifically between positions 50 and 54.[22,63] Analysis of the homeodomain sequences available through the NHGRI homeodomain resource[1] reveal that when position 50 is occupied by a glutamine residue (~63% of all known homeodomains), many different residues can occupy position 54, with the majority containing either alanine (~41%) or methionine (~40%) at this position. K50 homeodomains (~6% of listed homeodomains) have a stricter evolutionary requirement for position 54 (alanine 70%, glutamine 26%). A mutant form of the thyroid transcription factor with an unnatural combination of amino acid residues at positions 50 and 54 (Q50K, Y54M) fails to bind to the DNA site predicted by earlier studies of the individual contributions of these residues to DNA recognition (K50-TAATCC, M54-ATTAGG).[22] These results suggest that the stricter evolutionary covariance seen for K50 homeodomains is important for maintenance of DNA binding. Studies of the MATα2[26] and fushi tarazu[27] homeodomains also concluded that combinatorial effects from amino acid residues in positions 50 and 54 contribute to the specificity of a particular homeodomain. These analyses of evolutionary covariance highlight co-mingled functional and structural requirements for amino acid residues that cannot be approximated or dissected by analysis of individual point mutants like Q50K, without also considering intramolecular interactions in the unperturbed system.

Perhaps the most convincing argument for additional biophysical and thermodynamic studies of Bicoid is its unusual role as the only known protein that uses a homeodomain to regulate translation, as well as transcription, by binding to both DNA and RNA during early *Drosophila* development. Bicoid represses the translation of another *Drosophila* transcription factor, *caudal*, by binding to the 3′ untranslated region (UTR) of the *caudal* mRNA.[65–69] Both K50 and R54 are required for RNA recognition. Interestingly, of the 1063 unique homeodomain sequences currently listed, only 4.5% contain arginine in position 54 of the recognition helix,[1] and even more unusual is the Bicoid homeodomain, which is the only known homeodomain that contains a K50/R54 combination.[1] While a K50A mutation abolishes both DNA and RNA recognition by the homeodomain, mutation of R54 to alanine preferentially affects recognition of both RNA sequences,[70] and a non-consensus DNA site,[23] raising questions about the role of R54 in the discrimination of DNA/RNA recognition by the Bicoid homeodomain.

Despite numerous structural and functional analyses of homeodomains, such studies have not explained the effects of different combinations of amino acids on DNA-binding site preference and affinity. Crystallographic studies have generally indicated that there are several conserved and stable interactions at the protein–DNA interface, usually involving the nearly invariant N51.[52,58] This stable interaction is accompanied by multiple, significantly populated conformations for the side-chain of the

amino acid residue in position 50. On the other hand, NMR[36,45,71] and molecular dynamics simulations[37,71–73] have provided strong evidence for a dynamic, fluctuating environment, including motion of both the N51 and position 50 side-chains, and water in the interface. Billeter and co-workers suggest that a network of short-lived contacts during protein–DNA interaction reduces the entropic cost that would be incurred by more rigid side-chain–DNA interactions.[71] The study of the PITX2 K50 homeodomain has recently lent support to this theory by showing that the K50 side-chain resonances exhibit line broadening and multiple conformations, indicative of conformational flexibility.[36] This dichotomy of a dynamic and entropically favorable environment at the protein–DNA interface seen by NMR and molecular dynamics *versus* a more organized environment seen in X-ray studies presents an appealing frame of reference for the analysis of the solution structure of the Bicoid homeodomain, the effect of evolutionary covariance of amino acids on nucleic acid recognition, its dual role as both a transcriptional and translational regulatory protein, and the role of molecular dynamics in protein–DNA recognition.

To clarify the role of key amino acids during DNA recognition by Bicoid, the solution structure of the Bicoid homeodomain/consensus DNA binding site (TAATCC) complex was solved using NMR spectroscopy. This structure, combined with the solution structure of the PITX2 homeodomain that our group has solved recently,[36] provide the first analyses of native K50 homeodomain structures, and provide two distinct foundations for the future study of the kinetic, thermodynamic, and motional contributions of amino acids to both DNA and RNA recognition by homeodomains.

Table 1. Structural statistics

A. *Restraint statistics*	
Total restraints	1724
Total protein restraints	1244
Long range $\|i-j\|\geq 5$	133
Medium range $1<\|i-j\|<5$	250
Short range $\|i-j\|\leq 1$	693
Hydrogen bond[a]	34
Angle restraints (Φ and Ψ, ω)[a,b]	134
Total DNA restraints	447
Intrabase	126
Interbase	66
Watson–Crick	55
Angles (α,β,γ,δ,ε,ζ,χ)	200
Total protein–DNA	
Intermolecular restraints	33
B. *CYANA statistics*	
Target function (Å^2)(cycle1/cycle7)	12.29/0.49
Backbone RMSD(Å)(cycle1/cycle7)	2.94/0.97
C. *AMBER statistics (kcal/mol)*	
Mean AMBER energy	-7670 ± 53
Mean van der Waals energy	-874 ± 13
Mean electrostatic energy	-2150 ± 12
Mean RMSD from ideal	
Bond lengths (Å)	0.0072 ± 0.0005
Bond angles (deg.)	4.303 ± 0.06
Restraint violations	
Average per structure, >0.3 Å	9
Maximum restraint violation (Å)	<0.5
Angle violations	
Average per structure, >5 deg.	3
Maximum angle violation (deg.)	<10
D. *Mean RMS deviations*	
From mean structure (Å)	(bb/heavy atom)
Protein (residues 10–58)	$1.39\pm0.41/1.83\pm0.40$
DNA (w/o 5′/3′-terminal bases)	$2.09\pm0.58/1.89\pm0.50$
DNA (consensus site only)	$1.78\pm0.49/1.62\pm0.41$
Protein (10–58+consensus site)	$1.49\pm0.36/1.79\pm0.35$

The statistics listed here are indicative of the 20 structures deposited in the RCSB PDB (1ZQ3).

[a] Hydrogen bond and angle restraints were obtained *via* mutual agreement from three separate analyses: derivation of coupling constants from an HNHA experiment, TALOS chemical shift analysis, and the presence of i to $i-3$ or $i-4$ NOEs.

[b] The ω angle restraints were obtained *via* the AMBER program, and applied to all residues except the two proline, the N-terminal glycine, and the C-terminal serine residues.

Results and Discussion

Homeodomain structure determination

We have determined the structure of a 67 amino acid residue construct of the Bicoid homeodomain bound to a 13mer duplex DNA site containing the consensus 5′-TAATCC-3′/3′-ATTAGG-5′ site. The Bicoid homeodomain construct we used contains an additional N-terminal glycine (a product of the TEV cleavage site), and six extra C-terminal amino acid residues that belong to the native Bicoid protein (added to improve solubility and reduce aggregation) (Supplementary Data S2). As was the case for the PITX2 homeodomain, the Bicoid homeodomain precipitated out of solution if it was concentrated in the absence of the DNA; this is in contrast to the behavior of other homeodomains that have been studied under solution conditions, where millimolar concentrations of the free homeodomain have been reported. The K_D for the binding of the Bicoid homeodomain to the 13mer DNA site used in this study was measured to be $4.28(\pm0.26)\times10^{-10}$ M (see Supplementary Data S1). A total of 1724 restraints (Table 1) were used to calculate the structure of the Bicoid homeodomain–DNA complex, including 1076 protein distance restraints (133 long-range, 250 medium-range and 693 short range, ~22 restraints per amino acid residue) and 134 torsional restraints (φ, ψ, and ω angles) determined by consideration of the $^3J_{\alpha\text{-NH}}$ coupling constants from analysis of a 3D HNHA experiment, a C^α chemical shift analysis by the CYANA program, 34 hydrogen bond restraints predicted from i to $i+3$ and i to $i+4$ NOEs seen between backbone H^N protons, and a TALOS chemical shift analysis.[74] Seven cycles of torsion angle dynamics and simulated annealing were performed using CYANA 2.0[75] (Table 1B) to arrive at an ensemble of 20 structures that best satisfied the restraint data. In general, the CYANA calculation revealed that the Bicoid homeodomain adopts the three-helical global fold that has been seen in structures of other homeodomains (Figure 1(a)). The residues 10–21, 27–38, and 43–60 form helices 1,

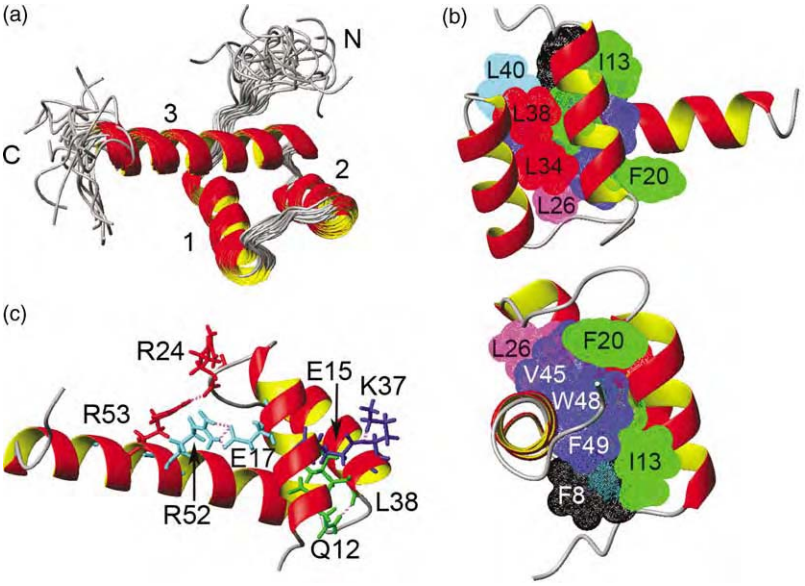

Figure 1. Structure of the Bicoid homeodomain before docking to the DNA, solved using CYANA2.0. (a) Ensemble of the 20 lowest energy conformers; the three helices and N and C termini are labeled. The second and third helices form the canonical helix-turn-helix protein motif. (b) Two views of the conserved amino acid residues of the hydrophobic core of the mean structure of the ensemble. The N terminus (F8, black), helix 1 (I13, L16, F20, green), the turn between helices 1 and 2 (L26, magenta), helix 2 (L34, L38, red), the turn between helices 2 and 3 (L40, cyan), and helix 3 (V45, W48, F49, blue) all contain amino acid residues that are important for formation of the hydrophobic core. (c) Long-distance hydrogen bonds/electrostatic interactions seen in the mean structure of the Bicoid homeodomain ensemble. Hydrogen bonds (pink dotted lines) were seen in a majority of conformers between Q12/L38 (green), E15/K37 (blue), E17/R52 (cyan), and R24/R53 (red).

2, and 3, respectively. The conserved (through amino acid identity or similarity) residues F8, I13, L16, F20, L26, L34, L38, L40, V45, W48, and F49 form the hydrophobic core of the homeodomain (Figure 1(b)), and provide interactions among the N-terminal arm (F8), helix 1 (I13, L16, F20), the turn between helices 1 and 2 (L26), helix 2 (L34, L38), the turn between helices 2 and 3 (L40) and helix 3 (V45, W48, F49), upon which the three-helical fold is formed. Helices 2 and 3 form the canonical helix-turn-helix DNA recognition motif that is present in homeodomains,[5] and was first seen in the crystal structures of the cI and Cro prokaryotic repressor proteins from bacteriophage lambda.[76,77] Additional stability for the homeodomain fold is provided by several long-range hydrogen bonds/electrostatic interactions seen in the ensemble of structures, including bonds between Q12 (terminal side-chain amide protons) /L38 (backbone carbonyl oxygen atom), E17 (terminal side-chain oxygen atom)/R52 (terminal side-chain amino protons), R24 (backbone carbonyl oxygen atom)/R53 (terminal amino protons), and E15 (terminal side-chain oxygen atom)/K37 (terminal side-chain amino protons) (Figure 1(c)).

Calculation of DNA structure

A schematic of the duplex DNA and the numbering system used here can be seen in Figure 2(a). Specific patterns of proton–proton distances seen in the 2D nuclear Overhauser effect spectroscopy (NOESY) spectrum of the DNA indicated that the duplex was adopting the *B*-DNA conformation (e.g. the short distances between the H2′ and H2″, and the H1′ and H5″, protons of sequential same-strand deoxyribose sugars).[78] This, in conjunction with previously determined NMR solution structures of homeodomain–DNA-binding sites shown to adopt the *B*-DNA conformation, allowed us to model our DNA sequence onto a general *B*-DNA structure using the subprogram NUCGEN[79] in the AMBER suite[80] (Figure 2(b), red). This model was subjected to 10 ps of NMR-refinement using 192 DNA proton–proton distance restraints, Watson–Crick restraints (to maintain base-pairing), and 200 angle restraints (standard *B*-DNA) using the AMBER all-atom force field with the Generalized Born solvation model,[81] resulting in a DNA structure (Figure 2(b), green) that reflects the DNA NMR-derived distance restraints. This DNA structure was then used during the docking calculations with the Bicoid homeodomain structure (Figure 2(b), blue).

Calculation of Bicoid homeodomain–DNA complex structure

A 2D $^{13}C(\omega_1)$-edited, $[^{13}C,^{15}N](\omega_2)$-filtered NOESY spectrum provided 33 unambiguous intermolecular distance restraints (Table 2; pertinent regions of the spectrum are available in Supplementary Data, Figure S3). Note that these 33 restraints describe interactions between protons separated by distances <6 Å, and do not necessarily indicate the presence of hydrophobic, electrostatic, or hydrogen bond interactions between the atoms. Additionally, these 33 NOEs do not necessarily describe all regions of close proximity between the protein and the DNA, due to technical limitations of the experiment, such as ambiguity in the assignment of unused NOEs and absence of possible NOE peaks due to limited sensitivity (e.g. exchange-broadened

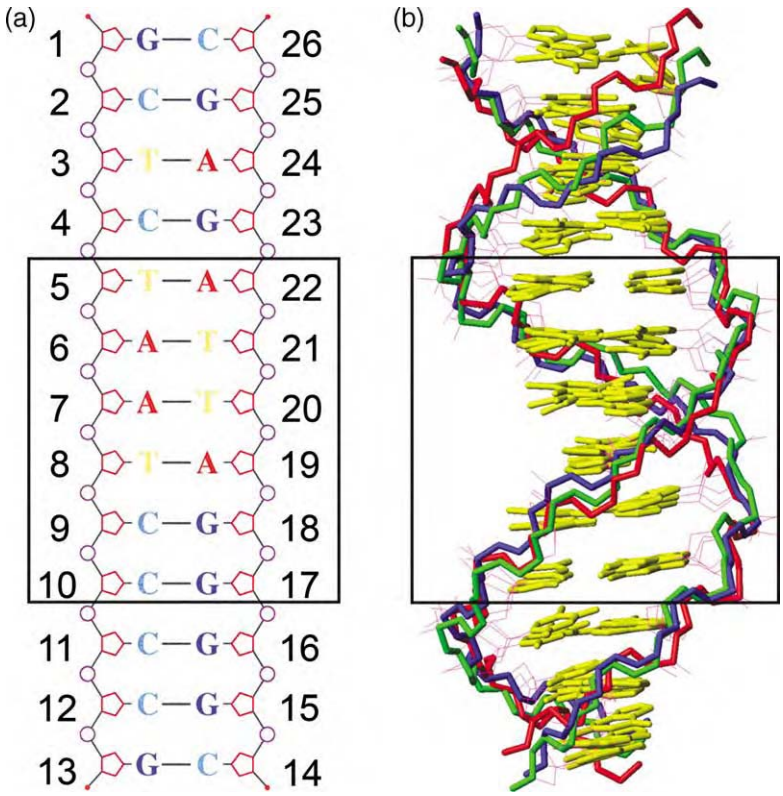

Figure 2. Description of the DNA site used for this study and its structure. (a) Schematic and numbering scheme of the 13mer DNA duplex used for this study. The 5′ ends of each strand are base numbers 1 and 14. (b) Three structures of the DNA seen during various stages of the structure calculation. The original *B*-DNA model created using the program NUCGEN (red), the mean structure after the initial docking calculation (green), and the mean structure after the final energy minimization in the absence of all constraints (blue) are superimposed.

peaks). Designations of strong, medium, and weak (indicated in parentheses in Table 2) were assigned to each NOE on the basis of peak volumes when possible, and by comparison and number of contour lines when peaks were overlapped, and were given upper distance limits of 4 Å, 5 Å, and 6 Å, respectively. Overall, five amino acid residues were in close proximity to the DNA, as detected by this NMR experiment, three of which are in helix 3 ("recognition helix") (I47, K50, and R54), one of which is located in the turn between helices 1 and 2 (Y25), and one of which is located in the N-terminal arm (R3). Nineteen of the 33 NOEs involved DNA bases in the TAATCC/ATTAGG consensus sequence, and ten of the 33 NOEs involved base protons (ex. H8 and H6).

All 1724 restraints described in Table 1 were used to calculate the docked protein–DNA structure as described in Materials and Methods. Briefly, the 20 protein structures provided by the CYANA program were placed 50 Å away from the NMR restraint-refined, double-stranded DNA duplex. The DNA was also rotated to achieve five different starting orientations relative to each of the 20 protein structures of the ensemble, which resulted in 100 different starting protein–DNA coordinate files which were input into the AMBER program. The resulting ensemble of docked structures (Figure 3) represents the 20 structures with the lowest violations of restraint data (Table 1C). Ramachandran analysis[82] *via* PROCHECK[83] indicated that 88.5% of the protein backbone residues were in the most favored conformational regions, with 10.3%, 0.7%, and 0.4% in the additionally allowed, generously allowed, and disallowed regions, respectively (see Supplementary Data S4).

Tertiary structure of the Bicoid homeodomain: comparison to other homeodomains

To better understand the unique multi-functionality of the Bicoid homeodomain, a comparative analysis of the backbone structures (C^{α} coordinates) of multiple homeodomains was performed. A structural alignment of 16 homeodomain structures, including the Bicoid homeodomain, was obtained using the MAMMOTH-mult server[84] at the Centro de Biología Molecular "Severo Ochoa" (CBMSO), which provides a rapid method to derive a superposition of the C^{α} coordinates of multiple input structures (see Materials and Methods for the homeodomains used; sequence alignment and an RMSD plot are available in Supplementary Data, Figure S5). Due to variability in the lengths of the N and C termini of homeodomains, structural alignment was performed on residues 8–55 of the homeodomains only, resulting in a pair-wise RMSD of 0.93 Å, confirming that the Bicoid homeodomain adopts an overall global fold similar to previously solved homeodomains structures. Analysis of the local RMSDs (RMSD of three amino acid segments) for the Bicoid homeodomain structure revealed that the positions of the C^{α} atoms of the amino acid residues at the end of helix 1 deviated from the other homeodomain structures by nearly 2.5 Å. The residues involved in this local deviation

Table 2. Protein–DNA intermolecular restraints

DNA base	DNA atom	Protein residue atom(s)	Protein
A6	H2″	I47	MD(m)[a]
A6	H2″	I47	MG(m)[b]
A7	H1′	I47	MD(m)
A7	H1′	I47	MG(m)
A7	H2″	I47	MD(m)
A7	H2″	I47	MG(m)
A7	H3′	I47	MD(s)
A7	H3′	I47	MG(m)
A7	H8	I47	MD(m)
A7	H8	I47	MG(s)
T8	H6	I47	MD(m)
T8	H6	I47	MG(m)
G16	H3′	Y25	QD(m)[c]
G16	H3′	Y25	QE(m)
G16	H4′	Y25	QD(s)
G16	H4′	Y25	QE(s)
G16	Q5′	Y25	QD(m)
G16	Q5′	Y25	QE(m)
G17	H3′	R54	QD(m)
G17	H4′	R54	QD(m)
G18	Q5′	R54	QD(m)
G17	H8	K50	HE2(w)
G17	H8	K50	HE3(w)
G17	H8	K50	HG2(w)
G17	H8	K50	HD2(m)
A22	H3′	R3	HD3(s)
A22	Q5′	R3	HD2(m)
A22	Q5′	R3	HD3(m)
A22	Q5′	R3	QG(m)
G23	Q5′	R3	HD2(s)
G23	Q5′	R3	HD3(s)
G23	Q5′	R3	QG(m)
G23	H8	R3	QG(m)

[a] Strength of the intermolecular NOEs are in parentheses: (s) 4 Å; (m) 5 Å; (w) 6 Å.
[b] MG and MD correspond to the two methyl groups of the isoleucine side-chain.
[c] For atoms that did not have a stereospecific assignment, restraints to pseudoatoms were used.

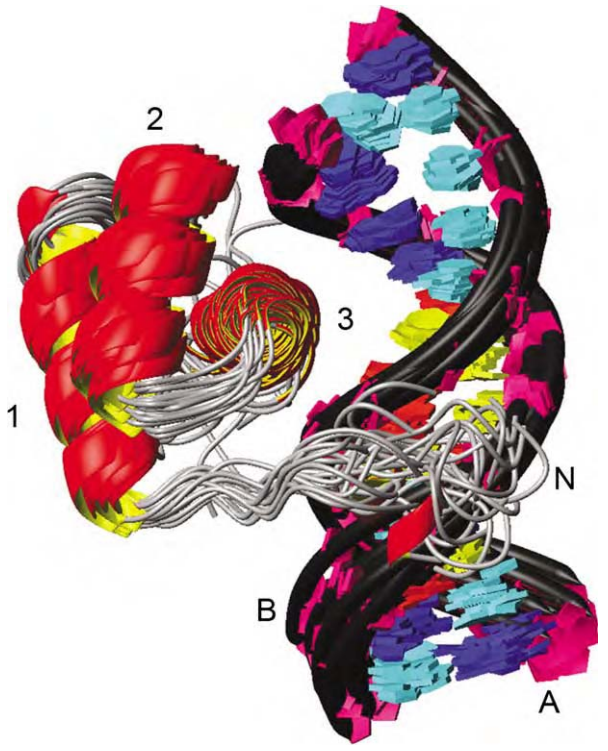

Figure 3. Superimposed ensemble of the 20 lowest energy structures from the AMBER docking calculation. Helices 1, 2, and 3, and the N terminus of the homeodomain are labeled. The sense and anti-sense strands are labeled A and B, respectively. Cytosine (cyan), guanosine (blue), thymine (yellow), and adenine (red) bases, deoxyribose sugars (magenta), and the phosphate backbone (black) are color-coded. Helix 3 inserts into the major groove and the N-terminal arm wraps around and contacts the minor groove. Statistics for the ensemble are given in Table 1.

include Q18, H19, F20, and L21 in helix 1, and residues Q22 and G23 in the turn between helices 1 and 2. A recent alanine scanning study of the engrailed homeodomain indicated that, while the side-chains of residues 18, 19, and 21 of this region were relatively unimportant for maintenance of engrailed function, residues 22 and 23 showed a strong preference for the wild-type residue, indicating the importance of side-chain identity for these residues in maintenance of DNA recognition by the homeodomain.[28] The requirement for phenylalanine in position 20 is well demonstrated by its conservation across the homeodomain family and its presence in the conserved hydrophobic core. However, the strong preferences for the E22 and N23 residues of the engrailed homeodomain were unexpected. To assess the relative importance of these residues to the homeodomain family, we analyzed the prevalence of amino acids at these positions across the 1063 known homeodomain sequences. Position 22 of the homeodomain showed very little evolutionary conservation, while position 23 showed a stronger preference for the asparagine residue (37%) observed in the engrailed study. Additionally, a clear preference for larger side-chains with hydrogen bonding capability can be seen for position 23 (see the Table in Supplementary Data, S6). Therefore, we hypothesize that the local RMSD differences seen in the Bicoid homeodomain structure are caused by the presence of a glycine residue in position 23 (seen in only 12/1063 known homeodomain sequences at this position, no other G23 homeodomain structures available), resulting in the alteration of the turn between helix 1 and helix 2, allowing helix 1 to move closer to helix 2, as seen in Figure 4.

Basis of DNA recognition by the Bicoid homeodomain

The intermolecular NOEs listed in Table 2, in combination with the intramolecular protein and DNA NOEs used to describe the structure of each component separately, describe (a) the global docking arrangement of the homeodomain on the DNA, and (b) the orientation of specific side-chains in the

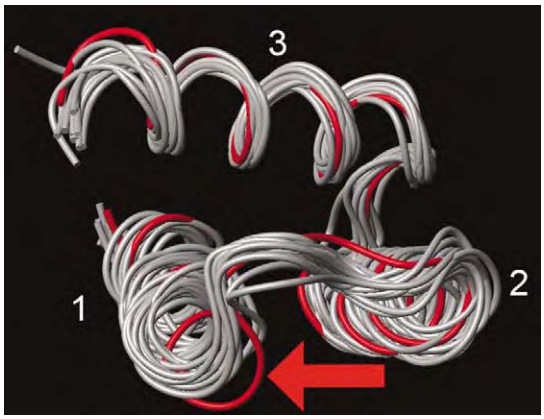

Figure 4. Superimposed C^{α} atom ribbon diagrams of 16 homeodomains (see the text for a list). The region of high local RMSD difference between the Bicoid homeodomain (red) and the rest (grey) is labeled with a red arrow. The Bicoid homeodomain contains glycine in position 23, which results in reduced hydrogen-bonding ability to nearby residues and an alteration of the homeodomain structure when compared to the 15 other homeodomain structures. The functional significance of this difference is unknown.

interface between the two. From these data alone, it is possible to draw some conclusions about specific side-chain–DNA contacts during recognition by the Bicoid homeodomain. However, a more extensive analysis of the basis of protein–DNA recognition, including water-mediated DNA recognition, can be achieved by using solvated molecular dynamics simulations, which approximate the energy of the system, including the effects of water molecules, salt concentrations, temperature, electrostatics, hydrophobic effects, etc. through the application of molecular and solvent force fields. The results of the AMBER docking calculation, in conjunction with the experimentally observed NOEs are discussed here.

Overall, three distinct regions of the homeodomain are involved in recognition of the DNA site: the N-terminal arm, the turn between helices 1 and 2, and the recognition helix. The N-terminal arm (residues 1–9), which has greater sequence variability among homeodomains than the more conserved helical regions (residues 10–58), has been shown in other homeodomain studies to be unstructured in the absence of DNA and to wrap around and make contacts in the minor groove when bound to DNA. This is corroborated in our Bicoid homeodomain–DNA structure by the presence of observed intermolecular NOEs (Table 2), and resultant position in the docked structure (Figure 3).

While interaction between Y25, located in the turn between helices 1 and 2, and the DNA (Figure 5(a)) is supported by the presence of six intermolecular NOEs to the sugar protons of base G16 (ATTAGGG), the functional significance of this residue is debatable. An engrailed homolog shotgun scanning study[28] showed that there was essentially no preference between tyrosine and phenylalanine for position 25 (F:Y, 1.7:1) and concluded that the primary role of an aromatic residue in position 25 is to form a π–cation interaction with R53. However, this conclusion is not necessarily supported by further analysis of the available homeodomain sequences, which reveals a predominance of tyrosine in position 25 (69%) compared to other amino acids (lysine 12%, arginine 5%, phenylalanine 1%), indicating that the function of Y25 in homeodomains may be a

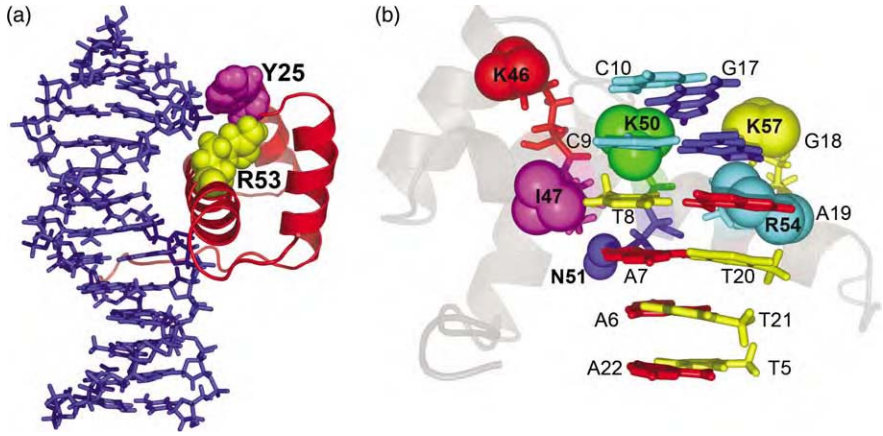

Figure 5. Interactions between the Bicoid homeodomain and the DNA. (a) The ribbon diagram of the Bicoid homeodomain (red) and the line diagram of the DNA (blue) are indicative of the mean structure of the protein–DNA complex. Six intermolecular NOEs placed Y25 (magenta) in close proximity to the DNA. A cation–π interaction with R53 (yellow) is seen in our structure, as well as other homeodomains. (b) Residues of the third helix involved in DNA recognition. The side-chains of K46 (red), I47 (magenta), K50 (green), N51 (blue), R54 (cyan), and K57 (yellow) are shown in relation to the protein backbone (grey). The bases of the recognition site are labeled (TAATCC/ATTAGG). Select atoms of the protein side-chains of interest (for lysine the terminal HZ1, 2, and 3 atoms, for isoleucine the HD1, 2, and 3 methyl protons, for N51 the HD21 and 22 atoms, and for R54 the terminal HH11, 12, 22, and 21 protons) are shown by spheres representing the RMSD of that atom from the mean, indicating the variability of those atoms in the ensemble.

more complex combination of hydrogen bonding and cation–π interactions with other amino acids of the homeodomain and bases of the DNA.[85] Another appealing, and untested, theory for the prevalence of tyrosine in position 25 is the possibility of phosphorylation,[86] which may affect DNA binding due to the repelling of like-negative charges of the phosphorylated tyrosine and the phosphate backbone of the DNA.

The most extensive region of the homeodomain involved in binding to the DNA site is the third "recognition" helix (Figure 5(b)). Interaction between three amino acids of the recognition helix (I47, K50, and R54) and the DNA were observed experimentally, providing 19 of the 33 detected intermolecular NOEs, and the majority of information used to describe the global homeodomain–DNA docking orientation and local orientation of these three side-chains in the protein–DNA interface. Interactions between I47 and the DNA were hydrophobic in nature, and involved close contacts between the methyl groups of the isoleucine side-chain and the methyl group of T8 and base/sugar protons of T8, A7, and A6 (TAATCC). K50 was involved in direct and water-mediated hydrogen bonds to bases of the sense and anti-sense strand of the consensus site (TAATCC/ATTAGG), and R54 was involved in direct and water-mediated hydrogen bonds to the G18 and A19 bases of the anti-sense strand of the consensus site (ATTAGG).

Crystallographic studies of homeodomain–DNA complexes invariably indicate that the N51 side-chain forms a pair of hydrogen bonds to the second adenine of the TAAT core, donating a hydrogen bond to the adenine N7 and accepting a hydrogen bond from the N6. While no intermolecular NOE was identified unambiguously between N51 and the DNA in our Bicoid data, the global positioning provided by the other observed intermolecular contacts places this side-chain in close proximity to A6 and A7 of the consensus site (TAATCC), as expected due to its conservation and previously studied role in DNA recognition by homeodomains. (A potential, weak intermolecular NOE between the N51 side-chain atom HD22 and the H8 proton of A7 was observed but not used in the calculations, due to ambiguity in assignment.) Very large, downfield chemical shifts were observed in both the ^{15}N and ^{1}H resonance frequencies for the terminal amide group of N51 (spectrum in Supplementary Data S7), which appears to be a conserved characteristic, as similar shifts have been observed in previous NMR studies of homeodomain–DNA complexes (PITX2,[36] Antennapedia,[87] and vnd/NK-2[45]). These unusually large shifts are presumably caused, at least in part, by the positioning of the terminal amide group near the plane of an aromatic base (ring current shifts) and by hydrogen bonding to a DNA base, most likely the second adenine of the TAAT core (TAAT), consistent with the conserved structural role of N51 during DNA recognition. In addition to the characteristic downfield chemical shifts, substantial line-broadening of the N51 side-chain amide proton resonances was observed in our NMR data, indicating the presence of microsecond–millisecond timescale fluctuations in the magnetic environment of the amide protons. The occurrence of such line-broadening, which could explain why no intermolecular NOE was assigned between the side-chain amide protons and the DNA, appears to be another conserved characteristic feature of N51 in homeodomain–DNA complexes; similar observations were reported in the PITX2,[36] Antennapedia,[87] and vnd/NK-2 studies.[45] The origin of the fluctuations in the magnetic environment of the N51 side-chain protons is currently unknown. Billeter *et al*. suggested that the lack of observed intermolecular NOEs involving the N51 side-chain and the presence of line-broadening could be due to a fluctuating network of weak interactions involving both adenine bases of the TAAT core and nearby water molecules.[35] On the other hand, Gruschus *et al*. did report the observation of intermolecular NOEs between the N51 side-chain amide protons and the DNA, possibly indicative of the more stable and specific interactions described in the crystallographic studies, and suggested that the line-broadening could be due to millisecond timescale motions of the side-chain and to strong interactions with nearby water molecules.[40] Based on their comparative study of the X-ray and NMR structures of the Antennapedia homeodomain–DNA complex, Fraenkel and Pabo hypothesized that several types of motions could lead to the observed line-broadening of the N51 side-chain resonances, and that such motions could still be consistent with the picture of the specific hydrogen bonding pattern seen in the X-ray data.[88] The suggested motions include occasional transitions to conformations significantly different from that observed in the X-ray structure, modest fluctuations in the χ2 angle of N51, and dynamic changes in the solvent structure around the N51 side-chain. Our Bicoid data are consistent with the model proposed by Fraenkel and Pabo.

Role of K50 and R54 in DNA recognition

Due to the ongoing discussion in the literature concerning the role of side-chain motion and water-mediated DNA recognition in the homeodomain family, and the role of R54 in discriminating between consensus and non-consensus DNA sites and RNA, we performed a more detailed analysis of the behavior of the K50 and R54 side-chains. Our analysis included the frequency and distribution of direct and water-mediated interactions between these side-chains and specific bases of the DNA. As indicated earlier, four inter-molecular NOEs between K50 and base G17 and three between R54 and bases G17 and G18 provided valuable information regarding the orientation of the recognition helix in the major groove of the DNA. However, because the majority of these NOEs were relatively weak, and consequently assigned a loose upper

Table 3. Direct and water-mediated K50–DNA hydrogen bonds

DNA base	DNA atom	Protein residue	Protein atom	Number in base	Direct or water ensemble specific?	Mediated
G17	O6	K50	QZ	1/20	Y	Water-mediated
G18	O6	K50	QZ	1/20	Y	Water-mediated
T8	O4	K50	QZ	2/20	Y	Direct
G18	O6	K50	QZ	5/20	Y	Direct
C9	O4	K50	QZ	5/20	Y	Water-mediated
G17	N7	K50	QZ	13/20	Y	Direct

limit of 6 Å, this allowed greater conformational flexibility for these side-chains during the docking calculation.

Analysis of the K50 side-chain reveals that 20 direct and eight water-mediated base-specific hydrogen bonds are seen in the 20 conformer ensemble, contacting four different bases of the consensus site (5′-TAATCC-3′/3′-ATTAGG-5′). The predominant interaction between the terminal amino protons of the K50 side-chain and the DNA are direct hydrogen bonds to the N7 atom of base G17 (ATTAGG, Table 3). This is consistent with the results seen in both the X-ray structure of the engrailed Q50K/DNA and the PITX2/DNA structures. This interaction was seen in 13 of 20 members of the ensemble, followed in frequency by direct hydrogen bonds to the O6 atom of G18 (ATTAGG, seen in five out of the 20 members of the ensemble), and water-mediated hydrogen bonds to the O4 atom of C9 (TAATCC, also five out of 20). Hydrogen bonds were sometimes bidentate, but were seen also in a variety of conformations involving a greater number of DNA bases of the consensus site than is possible from one or two significantly populated conformations, as suggested by the engrailed Q50K/DNA crystal structure (see Supplemental Data, Figure S8). These observations demonstrate that there are many conformations of the K50 side-chain that (a) satisfy intramolecular protein–protein NOEs involving K50 side-chain protons, (b) satisfy the observed intermolecular protein–DNA NOEs, and (c) satisfy the energetic parameters incorporated in the AMBER force field.

Additional evidence for multiple conformations of K50 is provided by the experimentally observed line broadening of the K50 side-chain resonances, evidence that was seen also for the side-chain of the PITX2 K50 residue, and for the side-chains in other protein–DNA complexes,[45,87,89,90] and can be indicative of side-chain motion on the intermediate timescale (*vide infra*). Line broadening of these resonances could be caused by many factors, including protein side-chain motion, DNA motion, and/or ring current effects from DNA bases or nearby aromatic residues. To address some of these possibilities, the behavior of the side-chain resonances of another lysine residue that contacts DNA, K46, was also examined. In Figure 6(a), resonances of the K46 and K50 side-chains, detected through their respective H^{ε} resonance frequencies, are compared (a comparison made easier by the unusual lack of significant spectral overlap for these H^{ε} resonances). Resonances of the K50 side-chain were very weakly detected in this HCCH-total correlated spectroscopy (TOCSY) experiment, as well as in the rest of the NMR experiments used in this study. The corresponding resonances of the K46 side-chain are, in comparison, strong and well-resolved. Line-broadening on the scale of that observed for the K50 side-chain was not seen for the DNA protons, suggesting that the line broadening was not caused by DNA motion. The variety of K50 conformations seen in the ensemble and the line broadening observed for the K50 side-chain resonances, while not indisputable evidence for K50 side-chain motion, taken together, are highly

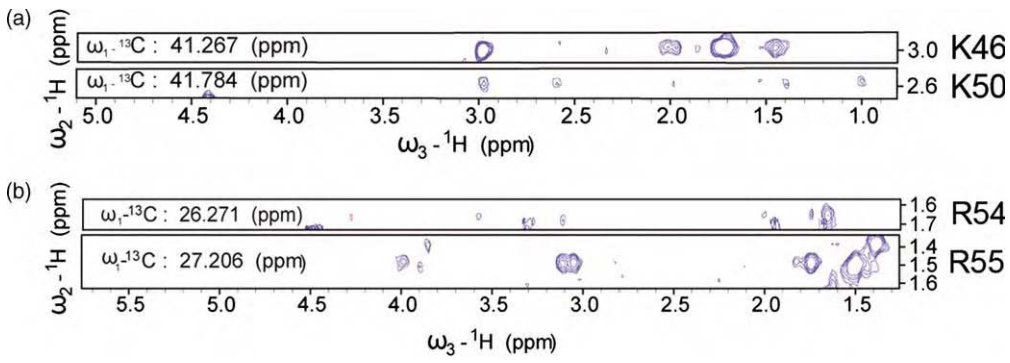

Figure 6. Line broadening of the K50 and R54 side-chain resonances. (a) The top panel shows the HCCH-TOCSY peaks of the K46 side-chain detected through the H^{ε} resonance compared to the same resonances of the K50 side-chain (bottom panel). (b) The top panel shows the HCCH-TOCSY peaks of the R54 side-chain detected through the H^{γ} resonance compared to the same resonances of the R55 side-chain (bottom panel).

Table 4. Direct and water-mediated R54–DNA hydrogen bonds

DNA base	DNA atom	Protein AA	Protein atom	Number in base	Direct or water ensemble specific?	Mediated
A19	H8	R54	QH1/2	1/20	Y	Water-mediated
G18	N7	R54	QH1/2	2/20	Y	Direct
A19	H62	R54	QH1/2	2/20	Y	Water-mediated
G18	O2P	R54	QH1/2	3/20	N	Water-mediated
G18	O2P	R54	QH1/2	5/20	N	Direct
G18	O2P	R54	HE	6/20	N	Water-mediated
A19	N7	R54	QH1/2	8/20	Y	Direct
G18	N7	R54	QH1/2	9/20	Y	Water-mediated
A19	N7	R54	QH1/2	13/20	Y	Water-mediated

suggestive of a dynamic role for K50 during recognition of DNA. Indications for side-chain dynamics of K50 were observed also in molecular dynamics simulations that we performed (see Conclusions).

The R54 side-chain, key to differentiation between consensus and non-consensus DNA and RNA-binding sites, was observed to make a greater number of hydrogen bonds to the DNA than K50 (K50:R54, 27:49). The N^{ε} group of R54 was shown to make six non-specific, water-mediated hydrogen bonds to the phosphate backbone of DNA base G18 (3′-ATTAGG-5′), leaving the two terminal amine groups to form the majority of base-specific contacts to the DNA. The predominant contact between R54 and the DNA observed in the ensemble was to the N7 atom of A19 (3′-ATTAGG-5′), both through direct (8/20) and water-mediated (13/20) hydrogen bonds (Table 4; see Supplementary Data Figure S9). The second most frequently seen base-specific contact involved the N7 atom of G18, to which direct (2/20) and water-mediated (9/20) hydrogen bonds were observed, followed in frequency by eight non-specific contacts to the phosphate backbone of G18 (direct:water-mediated, 5:3). The frequency and position of the R54/DNA interactions, similar to the previously discussed K50 side-chain, place the terminal amine groups in multiple conformations that support the possibility that this side-chain is in motion as well. Line broadening for R54 was examined and compared to resonances for R55, another amino acid in close proximity to the DNA (Figure 6(b)). Line broadening for this side-chain is observed also, although not as dramatically as that seen for K50, corroborating the possibility of motion on the intermediate timescale.

Conclusions

The results described in this study reveal several special features about the Bcd homeodomain structure. First, when compared with other homeodomains, the Bcd homeodomain exhibits a significant structural variation at the end of the first helix. Our analysis suggests that the unique glycine residue at position 23 may be responsible for this variation. Second, our study reveals evidence indicative of molecular motion of the side-chains of K50, N51, and R54, a finding that is supported also by preliminary results of solvated molecular dynamics simulations of the Bicoid homeodomain–DNA complex (our unpublished results). Importantly, both K50 and R54 play critical roles in recognizing both DNA and RNA, as evidenced by previous mutation analyses,[23,70,91] and our current structural data. On the consensus TAATCC DNA site, both side-chains make direct and water-mediated contacts to bases in the DNA (ATTAGG for R54, and TAATCC/ATTAGG for K50). Among all the homeodomains, Bcd is the only one that contains a K50/R54 combination and is currently the only homeodomain known to have the ability to recognize both DNA and RNA. We imagine that some of the structural features and conformational dynamics revealed in our current study may play important roles in Bcd recognition of RNA sequences as well as non-consensus DNA sites.

Materials and Methods

Preparation of the Bicoid homeodomain NMR sample

The 60 amino acid residue Bicoid homeodomain (plus seven residues C-terminal to the homeodomain, added for solubility reasons; the sequence is shown in Supplementary Data) was amplified from a full-length Bcd cDNA plasmid, pFY441,[92] for insertion into the pET41a(+) (Novagen) plasmid, using two primers. Primer A: 5′-GCAC**GAATTC**GAAAACCTGTATTTTCAGGGTCCACGTCGCACCCG-3′, and primer B: 5′-CGCGGC**AAGCTT**TTATTAGGACTGGTCCTTGTGCTGATCCG-3′†. Primer A contains an EcoRI site (bold), a 21 nucleotide sequence encoding the seven amino acid residue tobacco etch virus (TEV) protease cleavage site (underlined), and the first 14 nucleotides of the Bicoid homeodomain. Primer B contains the last 23 nucleotides of the Bicoid homeodomain, two stop codons (underlined), and a HindIII cleavage site (bold). Expression, purification, and preparation of the NMR sample were accomplished using modifications of procedures that have been described,[36] with the following exceptions: (1) after glutathione-*S*-transferase (GST) resin purification, the fusion protein was digested overnight at room temperature with 5 mg of active TEV protease (pRK793 plasmid obtained from Dr David Waugh at the National

† http://www.idtdna.com

Cancer Institute;[93] protein prepared in-house); (2) the resultant protein pool containing TEV and the Bicoid homeodomain was applied to 3 ml of equilibrated His-Bind resin (Novagen) to remove the His-tagged TEV protease, followed by an ion-exchange purification using 2 ml of equilibrated SP-Sepharose fast flow IEX resin (Amersham). From 1 l of doubly labeled rich minimal medium, typical yields of the Bicoid homeodomain were between 3 mg and 5 mg. Sense (5′-GCTCTAATCCCG-3′) and anti-sense (5′-CGGGGATTAGAGC-3′) strands of the DNA-binding site† were dissolved in nuclease-free water, mixed in an equimolar ratio, heated to 95 °C for 15 min, and allowed to cool to room temperature. A slight excess of the annealed DNA complex was added to the dialysis bag containing the purified Bicoid homeodomain. The protein/DNA solution was concentrated by the addition of SpectraGel absorbent (Spectrum) to the outside of the dialysis bag at 4 °C. The absorbent was reapplied until a final volume of 300–500 μl (0.75–1 mM complex) was reached. Sample was then dialyzed into 10 mM NaH_2PO_4 (pH 7.0), ^2H_2O was added to 10%, along with 1mM PMSF, 1 mM DTT, 1 mM EDTA, 0.1 mM NaN_3, 0.3 mM leupeptin, 0.2 mM Pefabloc (Roche), and dissolved crushed protease inhibitor tablets (Roche: one tablet dissolved in 3 ml H_2O, 1 μl added to 540 μl sample). Samples were placed into Shigemi NMR tubes and stored at 4 °C.

NMR spectroscopy and structure calculation

All experiments were performed with 600 MHz and 800 MHz Varian Inova spectrometers at 295 K and proton chemical shifts were referenced against an external DSS standard. Data were processed using the programs NMRDraw/NMRPipe.[94] Spectra were analyzed and assigned using the program SPARKY‡. The pulse programming codes were written in-house. Protein ^1H, ^{15}N, and ^{13}C resonance assignments (>95% complete, listed in Supplementary Data S10) were made using the following experiments: 2D ^{15}N heteronuclear single quantum coherence (HSQC), 3D HNCA, HNCO, CBCA(CO)NH, HN(CO)CA, and HNCACB,[95–99] 2D heteronuclear multiple quantum coherence (HMQC) and 2D HMQC-TOCSY,[100,101] 2D ^{13}C-HSQC, HBHA(CBCA-CO)NH, H(CCO)NH-TOCSY, HCCH-TOCSY, TOCSY-HSQC, four 3D-NOESY experiments: three ^{15}N-separated NOESY (τ=50 ms, 80 ms, 125 ms mixing times)[102] and one ^{13}C-separated NOESY (τ=150 ms).[98] Backbone φ dihedral angles were obtained by analysis of a 3D HNHA spectrum.[103] Unlabeled DNA proton resonances were assigned using three $^{13}C/^{15}N$-filtered 2D-NOESY experiments (τ=60 ms, 120 ms, 300 ms mixing times)[62] and two ω_2-filtered 2D TOCSY experiments (τ=42, 80 ms)[104] using a standard assignment strategy[78] (see chemical shift list in Supplementary Data S11). Distance restraints were obtained from NOEs as described.[105] NOEs between the Bicoid homeodomain and the DNA were identified in a 2D $^{13}C(\omega_1)$-edited, [^{13}C, ^{15}N](ω_2)-filtered NOESY spectrum.[104,106,107] The unassigned, integrated peak lists from two 3D-^{15}N NOESY (τ=50 ms, 80 ms) spectra and one 3D-^{13}C NOESY spectrum were both assigned and used by the program CYANA2.0 to calculate the structure of the homeodomain.[75,108] The 20 structures with the lowest target function (potential energy) were then used to calculate the structure of the Bicoid homeodomain–DNA complex using the program AMBER7.0.[80] The protein was docked onto the DNA structure using a modified version of a previously described protocol.[105] The 20 docked structures with the lowest energies and restraint violations were subjected to an additional 30 ps of conjugate-gradient energy minimization with solvent included using the SANDER module of AMBER 7.

Homeodomain comparison

The Bcd homeodomain C^α backbone structure (residues 8–55) was aligned with 15 other homeodomain structures (PDB IDs: 1YZ8 (PITX2),[36] 1AHD (Antennapedia),[35] 1FTT (thyroid transcription factor-1),[38] 1NK3 (Vnd/NK-2),[40] 1BW5 (Isl-1),[41] 1FTZ (fushi tarazu),[42] 1CQT (POU),[48] 1ENH (engrailed),[49] 1JGG (Even-skipped),[52] 1IG7 (Msx-1),[53] 1AU7 (Pit-1),[54] 1B8I (Ultrabithorax),[56] 1B72 (HoxB-1),[57] 2HDD (engrailed Q50K),[58] and 1FJL (Paired)[59]) using the MAMMOTH server§.[84] This list includes seven NMR[35,36,38,40–42] and nine X-ray[48,49,52–54,56,58,59] structures, 12 solved with DNA present and four without DNA.[38,41,42,49]

Data bank accession codes

The atomic coordinates have been deposited in the RCSB Protein Data Bank with accession code 1ZQ3. The protein chemical shift assignments have been submitted to the BMRB with accession code 6906.

Acknowledgements

This work was supported by a grant (GM063855) from the National Institutes of Health (to M.R.). Support from NSF and NIH to J.M. is also acknowledged. We thank Dr Jack Howarth for computer support and maintenance of the NMR facility, and Dr Carolyn Price for reagents, equipment, and advice during plasmid construction.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2005.12.007

References

1. Banerjee-Basu, S., Moreland, T., Hsu, B. J., Trout, K. L. & Baxevanis, A. D. (2003). The homeodomain resource: 2003 update. *Nucl. Acids Res.* **31**, 304–306.
2. Gehring, W. J., Affolter, M. & Burglin, T. (1994). Homeodomain proteins. *Annu. Rev. Biochem.* **63**, 487–526.
3. Gehring, W. J., Qian, Y. Q., Billeter, M., Furukubo-Tokunaga, K., Schier, A. F., Resendez-Perez, D. *et al.* (1994). Homeodomain-DNA recognition. *Cell*, **78**, 211–223.
4. McGinnis, W. & Krumlauf, R. (1992). Homeobox genes and axial patterning. *Cell*, **68**, 283–302.

‡ http://www.cgl.ucsf.edu/home/sparky/

§ http://ub.cbm.uam.es/mammoth/mult

5. Treisman, J., Harris, E., Wilson, D. & Desplan, C. (1992). The homeodomain: a new face for the helix-turn-helix? *BioEssays*, **14**, 145–150.
6. Wolberger, C. (1993). Structure and DNA binding of the yeast MAT alpha 2 homeodomain. *Cold Spring Harbor Symp. Quant. Biol.* **58**, 159–166.
7. Hadrys, T., DeSalle, R., Sagasser, S., Fischer, N. & Schierwater, B. (2005). The Trichoplax PaxB gene: a putative Proto-PaxA/B/C gene predating the origin of nerve and sensory cells. *Mol. Biol. Evol.* **22**, 1569–1578.
8. Chambers, I. (2004). The molecular basis of pluripotency in mouse embryonic stem cells. *Cloning Stem Cells*, **6**, 386–391.
9. Berleth, T., Burri, M., Thoma, G., Bopp, D., Richstein, S., Frigerio, G. *et al*. (1988). The role of localization of bicoid RNA in organizing the anterior pattern of the *Drosophila* embryo. *EMBO J.* **7**, 1749–1756.
10. Driever, W. (2004). The bicoid morphogen papers (II): account from Wolfgang Driever. *Cell*, **116**, S7–S9.
11. Rivera-Pomar, R., Lu, X., Perrimon, N., Taubert, H. & Jackle, H. (1995). Activation of posterior gap gene expression in the *Drosophila* blastoderm. *Nature*, **376**, 253–256.
12. Struhl, G., Struhl, K. & Macdonald, P. M. (1989). The gradient morphogen bicoid is a concentration-dependent transcriptional activator. *Cell*, **57**, 1259–1273.
13. Burz, D. S., Rivera-Pomar, R., Jackle, H. & Hanes, S. D. (1998). Cooperative DNA-binding by Bicoid provides a mechanism for threshold-dependent gene activation in the *Drosophila* embryo. *EMBO J.* **17**, 5998–6009.
14. Gao, Q. & Finkelstein, R. (1998). Targeting gene expression to the head: the *Drosophila* orthodenticle gene is a direct target of the Bicoid morphogen. *Development*, **125**, 4185–4193.
15. Hanes, S. D. & Brent, R. (1989). DNA specificity of the bicoid activator protein is determined by homeodomain recognition helix residue 9. *Cell*, **57**, 1275–1283.
16. La Rosee, A., Hader, T., Taubert, H., Rivera-Pomar, R. & Jackle, H. (1997). Mechanism and Bicoid-dependent control of hairy stripe 7 expression in the posterior region of the *Drosophila* embryo. *EMBO J.* **16**, 4403–4411.
17. Wilson, D. S., Sheng, G., Jun, S. & Desplan, C. (1996). Conservation and diversification in homeodomain–DNA interactions: a comparative genetic analysis. *Proc. Natl Acad. Sci. USA*, **93**, 6886–6891.
18. Yuan, D., Ma, X. & Ma, J. (1996). Sequences outside the homeodomain of bicoid are required for protein-protein interaction. *J. Biol. Chem.* **271**, 21660–21665.
19. Ades, S. E. & Sauer, R. T. (1994). Differential DNA-binding specificity of the engrailed homeodomain: the role of residue 50. *Biochemistry,* **33**, 9187–9194.
20. Affolter, M., Percival-Smith, A., Muller, M., Leupin, W. & Gehring, W. J. (1990). DNA binding properties of the purified Antennapedia homeodomain. *Proc. Natl Acad. Sci. USA*, **87**, 4093–4097.
21. Damante, G., Fabbro, D., Pellizzari, L., Civitareale, D., Guazzi, S., Polycarpou-Schwartz, M. *et al*. (1994). Sequence-specific DNA recognition by the thyroid transcription factor-1 homeodomain. *Nucl. Acids Res.* **22**, 3075–3083.
22. Damante, G., Pellizzari, L., Esposito, G., Fogolari, F., Viglino, P., Fabbro, D. *et al*. (1996). A molecular code dictates sequence-specific DNA recognition by homeodomains. *EMBO J.* **15**, 4992–5000.
23. Dave, V., Zhao, C., Yang, F., Tung, C. S. & Ma, J. (2000). Reprogrammable recognition codes in bicoid homeodomain–DNA interaction. *Mol. Cell. Biol.* **20**, 7673–7684.
24. Florence, B., Handrow, R. & Laughon, A. (1991). DNA-binding specificity of the fushi tarazu homeodomain. *Mol. Cel.l Biol.* **11**, 3613–3623.
25. Hanes, S. D. & Brent, R. (1991). A genetic model for interaction of the homeodomain recognition helix with DNA. *Science*, **251**, 426–430.
26. Mathias, J. R., Zhong, H., Jin, Y. & Vershon, A. K. (2001). Altering the DNA-binding specificity of the yeast Matalpha 2 homeodomain protein. *J. Biol. Chem.* **276**, 32696–32703.
27. Percival-Smith, A., Muller, M., Affolter, M. & Gehring, W. J. (1990). The interaction with DNA of wild-type and mutant fushi tarazu homeodomains. *EMBO J.* **9**, 3967–3974.
28. Sato, K., Simon, M. D., Levin, A. M., Shokat, K. M. & Weiss, G. A. (2004). Dissecting the Engrailed homeodomain-DNA interaction by phage-displayed shotgun scanning. *Chem. Biol.* **11**, 1017–1023.
29. Schier, A. F. & Gehring, W. J. (1993). Functional specificity of the homeodomain protein fushi tarazu: the role of DNA-binding specificity *in vivo*. *Proc. Natl Acad. Sci. USA*, **90**, 1450–1454.
30. Treisman, J., Gonczy, P., Vashishtha, M., Harris, E. & Desplan, C. (1989). A single amino acid can determine the DNA binding specificity of homeodomain proteins. *Cell*, **59**, 553–562.
31. Weiler, S., Gruschus, J. M., Tsao, D. H., Yu, L., Wang, L. H., Nirenberg, M. & Ferretti, J. A. (1998). Site-directed mutations in the vnd/NK-2 homeodomain. Basis of variations in structure and sequence-specific DNA binding. *J. Biol. Chem.* **273**, 10994–11000.
32. Yuan, D., Ma, X. & Ma, J. (1999). Recognition of multiple patterns of DNA sites by *Drosophila* homeodomain protein Bicoid. *J. Biochem. (Tokyo)*, **125**, 809–817.
33. Zhao, C., Dave, V., Yang, F., Scarborough, T. & Ma, J. (2000). Target selectivity of bicoid is dependent on nonconsensus site recognition and protein-protein interaction. *Mol. Cell. Biol.* **20**, 8112–8123.
34. Anderson, J. S., Forman, M. D., Modleski, S., Dahlquist, F. W. & Baxter, S. M. (2000). Cooperative ordering in homeodomain-DNA recognition: solution structure and dynamics of the MATa1 homeodomain. *Biochemistry*, **39**, 10045–11054.
35. Billeter, M., Qian, Y. Q., Otting, G., Müller, M., Gehring, W. & Wüthrich, K. (1993). Determination of the nuclear magnetic resonance solution structure of an Antennapedia homeodomain–DNA complex. *J. Mol. Biol.* **234**, 1084–1093.
36. Chaney, B. A., Clark-Baldwin, K., Dave, V., Ma, J. & Rance, M. (2005). Solution structure of the K50 class homeodomain PITX2 bound to DNA and implications for mutations that cause Rieger syndrome. *Biochemistry*, **44**, 7497–7511.
37. Cox, M., van Tilborg, P. J., de Laat, W., Boelens, R., van Leeuwen, H. C., van der Vliet, P. C. & Kaptein, R. (1995). Solution structure of the Oct-1 POU homeodomain determined by NMR and restrained molecular dynamics. *J. Biomol. NMR*, **6**, 23–32.
38. Esposito, G., Fogolari, F., Damante, G., Formisano, S., Tell, G., Leonardi, A. *et al*. (1996). Analysis of the solution structure of the homeodomain of rat thyroid

transcription factor 1 by 1H-NMR spectroscopy and restrained molecular mechanics. *Eur. J. Biochem.* **241**, 101–113.
39. Fernandez, C., Szyperski, T., Ono, A., Iwai, H., Tate, S., Kainosho, M. & Wüthrich, K. (1998). NMR with 13C, 15N-doubly-labeled DNA: the Antennapedia homeodomain complex with a 14-mer DNA duplex. *J. Biomol. NMR*, **12**, 25–37.
40. Gruschus, J. M., Tsao, D. H., Wang, L. H., Nirenberg, M. & Ferretti, J. A. (1999). The three-dimensional structure of the vnd/NK-2 homeodomain-DNA complex by NMR spectroscopy. *J. Mol. Biol.* **289**, 529–545.
41. Ippel, H., Larsson, G., Behravan, G., Zdunek, J., Lundqvist, M., Schleucher, J. *et al*. (1999). The solution structure of the homeodomain of the rat insulin-gene enhancer protein isl-1. Comparison with other homeodomains. *J. Mol. Biol.* **288**, 689–703.
42. Qian, Y. Q., Furukubo-Tokunaga, K., Resendez-Perez, D., Müller, M., Gehring, W. J. & Wüthrich, K. (1994). Nuclear magnetic resonance solution structure of the fushi tarazu homeodomain from *Drosophila* and comparison with the Antennapedia homeodomain. *J. Mol. Biol.* **238**, 333–345.
43. Schott, O., Billeter, M., Leiting, B., Wider, G. & Wüthrich, K. (1997). The NMR solution structure of the non-classical homeodomain from the rat liver LFB1/HNF1 transcription factor. *J. Mol. Biol.* **267**, 673–683.
44. Sivaraja, M., Botfield, M. C., Mueller, M., Jancso, A. & Weiss, M. A. (1994). Solution structure of a POU-specific homeodomain: 3D-NMR studies of human B-cell transcription factor Oct-2. *Biochemistry*, **33**, 9845–9855.
45. Tsao, D. H., Gruschus, J. M., Wang, L. H., Nirenberg, M. & Ferretti, J. A. (1995). The three-dimensional solution structure of the NK-2 homeodomain from *Drosophila*. *J. Mol. Biol.* **251**, 297–307.
46. Qian, Y. Q., Billeter, M., Otting, G., Müller, M., Gehring, W. J. & Wüthrich, K. (1989). The structure of the Antennapedia homeodomain determined by NMR spectroscopy in solution: comparison with prokaryotic repressors. *Cell*, **59**, 573–580.
47. Aishima, J. & Wolberger, C. (2003). Insights into nonspecific binding of homeodomains from a structure of MATalpha2 bound to DNA. *Proteins: Struct. Funct. Genet.* **51**, 544–551.
48. Chasman, D., Cepek, K., Sharp, P. A. & Pabo, C. O. (1999). Crystal structure of an OCA-B peptide bound to an Oct-1 POU domain/octamer DNA complex: specific recognition of a protein–DNA interface. *Genes Dev.* **13**, 2650–2657.
49. Clarke, N. D., Kissinger, C. R., Desjarlais, J., Gilliland, G. L. & Pabo, C. O. (1994). Structural studies of the engrailed homeodomain. *Protein Sci.* **3**, 1779–1787.
50. Grant, R. A., Rould, M. A., Klemm, J. D. & Pabo, C. O. (2000). Exploring the role of glutamine 50 in the homeodomain–DNA interface: crystal structure of engrailed (Gln50(Ala) complex at 2.0 Å. *Biochemistry*, **39**, 8187–8192.
51. Fraenkel, E., Rould, M. A., Chambers, K. A. & Pabo, C. O. (1998). Engrailed homeodomain-DNA complex at 2.2 Å resolution: a detailed view of the interface and comparison with other engrailed structures. *J. Mol. Biol.* **284**, 351–361.
52. Hirsch, J. A. & Aggarwal, A. K. (1995). Structure of the even-skipped homeodomain complexed to AT-rich DNA: new perspectives on homeodomain specificity. *EMBO J.* **14**, 6280–6291.
53. Hovde, S., Abate-Shen, C. & Geiger, J. H. (2001). Crystal structure of the Msx-1 homeodomain/DNA complex. *Biochemistry*, **40**, 12013–12021.
54. Jacobson, E. M., Li, P., Leon-del-Rio, A., Rosenfeld, M. G. & Aggarwal, A. K. (1997). Structure of Pit-1 POU domain bound to DNA as a dimer: unexpected arrangement and flexibility. *Genes Dev.* **11**, 198–212.
55. Kissinger, C. R., Liu, B. S., Martin-Blanco, E., Kornberg, T. B. & Pabo, C. O. (1990). Crystal structure of an engrailed homeodomain–DNA complex at 2.8 Å resolution: a framework for understanding homeodomain–DNA interactions. *Cell*, **63**, 579–590.
56. Passner, J. M., Ryoo, H. D., Shen, L., Mann, R. S. & Aggarwal, A. K. (1999). Structure of a DNA-bound Ultrabithorax-Extradenticle homeodomain complex. *Nature*, **397**, 714–719.
57. Piper, D. E., Batchelor, A. H., Chang, C. P., Cleary, M. L. & Wolberger, C. (1999). Structure of a HoxB1-Pbx1 heterodimer bound to DNA: role of the hexapeptide and a fourth homeodomain helix in complex formation. *Cell*, **96**, 587–597.
58. Tucker-Kellogg, L., Rould, M. A., Chambers, K. A., Ades, S. E., Sauer, R. T. & Pabo, C. O. (1997). Engrailed (Gln50(Lys) homeodomain-DNA complex at 1.9 Å resolution: structural basis for enhanced affinity and altered specificity. *Structure*, **5**, 1047–1054.
59. Wilson, D. S., Guenther, B., Desplan, C. & Kuriyan, J. (1995). High resolution crystal structure of a paired (Pax) class cooperative homeodomain dimer on DNA. *Cell*, **82**, 709–719.
60. Wolberger, C., Vershon, A. K., Liu, B., Johnson, A. D. & Pabo, C. O. (1991). Crystal structure of a MAT alpha 2 homeodomain-operator complex suggests a general model for homeodomain–DNA interactions. *Cell*, **67**, 517–528.
61. Laughon, A. (1991). DNA binding specificity of homeodomains. *Biochemistry*, **30**, 11357–11367.
62. Otting, G. & Wüthrich, K. (1990). Heteronuclear filters in two-dimensional [1H,1H]-NMR spectroscopy: combined use with isotope labelling for studies of macromolecular conformation and intermolecular interactions. *Quart. Rev. Biophys.* **23**, 39–96.
63. Pellizzari, L., Tell, G., Fabbro, D., Pucillo, C. & Damante, G. (1997). Functional interference between contacting amino acids of homeodomains. *FEBS Letters*, **407**, 320–324.
64. Billeter, M. (1996). Homeodomain-type DNA recognition. *Prog. Biophys. Mol. Biol.* **66**, 211–225.
65. Chan, S. K. & Struhl, G. (1997). Sequence-specific RNA binding by bicoid. *Nature*, **388**, 634.
66. Cho, P. F., Poulin, F., Cho-Park, Y. A., Cho-Park, I. B., Chicoine, J. D., Lasko, P. & Sonenberg, N. (2005). A new paradigm for translational control: inhibition *via* 5′-3′ mRNA tethering by Bicoid and the eIF4E cognate 4EHP. *Cell*, **121**, 411–423.
67. Dubnau, J. & Struhl, G. (1996). RNA recognition and translational regulation by a homeodomain protein. *Nature*, **379**, 694–699.
68. Niessing, D., Blanke, S. & Jackle, H. (2002). Bicoid associates with the 5′-cap-bound complex of caudal mRNA and represses translation. *Genes Dev.* **16**, 2576–2582.
69. Rivera-Pomar, R., Niessing, D., Schmidt-Ott, U., Gehring, W. J. & Jackle, H. (1996). RNA binding

and translational suppression by bicoid. *Nature*, **379**, 746–749.
70. Niessing, D., Driever, W., Sprenger, F., Taubert, H., Jackle, H. & Rivera-Pomar, R. (2000). Homeodomain position 54 specifies transcriptional versus translational control by Bicoid. *Mol. Cell.* **5**, 395–401.
71. Billeter, M., Güntert, P., Luginbühl, P. & Wüthrich, K. (1996). Hydration and DNA recognition by homeodomains. *Cell*, **85**, 1057–1065.
72. Gutmanas, A. & Billeter, M. (2004). Specific DNA recognition by the Antp homeodomain: MD simulations of specific and nonspecific complexes. *Proteins: Struct. Funct. Genet.* **57**, 772–782.
73. Li, H., Tejero, R., Monleon, D., Bassolino-Klimas, D., Abate-Shen, C., Bruccoleri, R. E. & Montelione, G. T. (1997). Homology modeling using simulated annealing of restrained molecular dynamics and conformational search calculations with CONGEN: application in predicting the three-dimensional structure of murine homeodomain Msx-1. *Protein Sci.* **6**, 956–970.
74. Cornilescu, G., Delaglio, F. & Bax, A. (1999). Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR*, **13**, 289–302.
75. Guntert, P. (2004). Automated NMR structure calculation with CYANA. *Methods Mol. Biol.* **278**, 353–378.
76. Anderson, W. F., Ohlendorf, D. H., Takeda, Y. & Matthews, B. W. (1981). Structure of the cro repressor from bacteriophage lambda and its interaction with DNA. *Nature*, **290**, 754–758.
77. Pabo, C. O. & Lewis, M. (1982). The operator-binding domain of lambda repressor: structure and DNA recognition. *Nature*, **298**, 443–447.
78. Wüthrich, K. (1986). *NMR of Proteins and Nucleic Acids*, Wiley & Sons, New York.
79. Bansal, M., Bhattacharyya, D. & Ravi, B. (1995). NUPARM and NUCGEN: software for analysis and generation of sequence dependent nucleic acid structures. *Comput. Appl. Biosci.* **11**, 281–287.
80. Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. S., Cheatham, T. E., deBolt, S. *et al*. (1995). AMBER, a computer program for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to elucidate the structures and energies of molecules. *Comput. Phys. Commun.* **91**, 1–41.
81. Onufriev, A., Case, D. A. & Bashford, D. (2002). Effective Born radii in the generalized Born approximation: the importance of being perfect. *J. Comput. Chem.* **23**, 1297–1304.
82. Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. (1963). Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* **7**, 95–99.
83. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283–291.
84. Ortiz, A. R., Strauss, C. E. & Olmea, O. (2002). MAMMOTH (matching molecular models obtained from theory): an automated method for model comparison. *Protein Sci.* **11**, 2606–2621.
85. Rooman, M., Lievin, J., Buisine, E. & Wintjens, R. (2002). Cation-pi/H-bond stair motifs at protein–DNA interfaces. *J. Mol. Biol.* **319**, 67–76.
86. Wolfe, S. A. (2004). Mapping key elements of a protein motif. *Chem. Biol.* **11**, 889–891.
87. Qian, Y. Q., Otting, G., Billeter, M., Müller, M., Gehring, W. & Wüthrich, K. (1993). Nuclear magnetic resonance spectroscopy of a DNA complex with the uniformly 13C-labeled Antennapedia homeodomain and structure determination of the DNA-bound homeodomain. *J. Mol. Biol.* **234**, 1070–1083.
88. Fraenkel, E. & Pabo, C. O. (1998). Comparison of X-ray and NMR structures for the Antennapedia homeodomain-DNA complex. *Nature Struct. Biol.* **5**, 692–697.
89. Foster, M. P., Wuttke, D. S., Radhakrishnan, I., Case, D. A., Gottesfeld, J. M. & Wright, P. E. (1997). Domain packing and dynamics in the DNA complex of the N-terminal zinc fingers of TFIIIA. *Nature Struct. Biol.* **4**, 605–608.
90. Nishikawa, T., Okamura, H., Nagadoi, A., Konig, P., Rhodes, D. & Nishimura, Y. (2001). Solution structure of a telomeric DNA complex of human TRF1. *Structure (Camb)*, **9**, 1237–1251.
91. Hanes, S. D., Riddihough, G., Ish-Horowicz, D. & Brent, R. (1994). Specific DNA recognition and intersite spacing are critical for action of the bicoid morphogen. *Mol. Cell. Biol.* **14**, 3364–3375.
92. Zhao, C., York, A., Yang, F., Forsthoefel, D. J., Dave, V., Fu, D. *et al*. (2002). The activity of the *Drosophila* morphogenetic protein Bicoid is inhibited by a domain located outside its homeodomain. *Development*, **129**, 1669–1680.
93. Kapust, R. B., Tozer, J., Anderson, D. E., Cherry, S., Copelend, T. D. & Waugh, D. S. (2001). Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. *Protein Eng.* **12**, 993–1000.
94. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, **6**, 277–293.
95. Grzesiek, S. & Bax, A. (1992). Improved 3D triple-resonance NMR techniques applied to a 31-KDa protein. *J. Magn. Reson.* **96**, 432–440.
96. Grzesiek, S. & Bax, A. (1992). Correlating backbone amide and side-chain resonances in larger proteins by multiple relayed triple resonance NMR. *J. Am. Chem. Soc.* **114**, 6291–6293.
97. Kay, L. E., Xu, G. Y. & Yamazaki, T. (1994). Enhanced-sensitivity triple resonance spectroscopy with minimal H_2O saturation. *J. Magn. Reson. ser. A*, **109**, 129–133.
98. Sattler, M., Schleucher, J. & Griesinger, C. (1999). Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. NMR Spectrosc.* **34**, 93–158.
99. Wittekind, M. & Mueller, L. (1993). HNCACB, a high-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the alpha carbon and beta carbon resonances in proteins. *J. Magn. Reson. ser. B*, **101**, 201–205.
100. Marion, D., Driscoll, P. C., Kay, L. E., Wingfield, P. T., Bax, A., Gronenborn, A. M. & Clore, G. M. (1989). Overcoming the overlap problem in the assignment of 1H NMR spectra of larger proteins by use of three-dimensional heteronuclear 1H-15N Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1 beta. *Biochemistry*, **28**, 6150–6156.
101. Zerbe, O., Szyperski, T., Ottiger, M. & Wüthrich, K. (1996). Three-dimensional ^{1}H-TOCSY-relayed

ct-^{13}C,^{1}H-HMQC for aromatic spin system identification in uniformly ^{13}C-labeled proteins. *J. Biomol. NMR*, **7**, 99–106.
102. Talluri, S. & Wagner, G. (1996). An optimized 3D NOESY-HSQC. *J. Magn. Reson. ser. B*, **112**, 200–205.
103. Vuister, G. W. & Bax, A. (1993). Quantitative J correlations: a new approach for measuring homonuclear three-bond J(HNHα) coupling constants in ^{15}N-enriched proteins. *J. Am. Chem. Soc.* **115**, 7772–7777.
104. Breeze, A. L. (2000). Isotope-filtered NMR methods for the study of biomolecular structure and interactions. *Prog. NMR Spectrosc.* **36**, 323–372.
105. Wuttke, D. S., Foster, M. P., Case, D. A., Gottesfeld, J. M. & Wright, P. E. (1997). Solution structure of the first three zinc fingers of TFIIIA bound to the cognate DNA sequence: determinants of affinity and sequence specificity. *J. Mol. Biol.* **273**, 183–206.
106. Lee, W., Revington, M. J., Arrowsmith, C. & Kay, L. E. (1994). A pulsed field gradient isotope-filtered 3D ^{13}C HMQC-NOESY experiment for extracting intermolecular NOE contacts in molecular complexes. *FEBS Letters*, **350**, 87–90.
107. Stuart, A. C., Borzilleri, K. A., Withka, J. M. & Palmer, A. G. (1999). Compensation for variations in ^{1}H–^{13}C scalar coupling constants in isotope-filtered NMR experiments. *J. Am. Chem. Soc.* **121**, 5346–5347.
108. Herrmann, T., Güntert, P. & Wüthrich, K. (2002). Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol.* **319**, 209–227.

Edited by P. Wright

(Received 11 September 2005; received in revised form 30 November 2005; accepted 2 December 2005)
Available online 22 December 2005

Review TRENDS in Genetics Vol.21 No.1 January 2005

Crossing the line between activation and repression

Jun Ma

Division of Developmental Biology, Cincinnati Children's Hospital Research Foundation, University of Cincinnati College of Medicine, 3333 Burnet Avenue, Cincinnati, OH 45229, USA

Gene transcription can be activated or repressed. Such seemingly simple decisions reflect the coordinated actions of a wide array of proteins. Activators and co-activators work together to stimulate the assembly and activity of the machinery that transcribes the gene, whereas repressors and co-repressors work to achieve the opposite goal. Recent studies show that many proteins often engage in regulatory activities and interactions that cross the activation-repression divide. This article discusses selected examples to illustrate the dynamic nature of the transcriptional regulation process and highlights the important roles of not only the individual proteins but also their communication system.

Transcription is the first step in expressing the genetic information of a cell. It is a highly regulated process that requires the coordinated actions of many different proteins [1,2]. For RNA polymerase II transcription, such factors can be loosely categorized into three broad groups. (i) General transcription factors (GTFs): these proteins, together with the RNA polymerase (RNAP), assemble into the transcription machinery at promoters. Although the majority of GTFs do not recognize specific DNA sequences, the TATA-binding protein (TBP) can bind to the TATA box in a promoter directly; other promoter elements such as the initiator element (INR) can provide additional specificity, particularly for TATA-less promoters. (ii) DNA-binding regulatory proteins, which are simply referred to as transcription factors (TFs) in this article. These proteins specifically bind to regulatory DNA sequences near – or sometimes distant from – gene promoters. These regulatory sequences include all DNA elements that can influence gene transcription such as enhancers, silencers, upstream activating sequences (UASs) and upstream repressing sequences (URSs). Regulatory elements tend to contain arrays of binding sites for TFs and the specific arrangements of these sites can affect how the factors work with one another. (iii) Co-factors: these proteins generally do not bind to specific DNA sequences themselves but can interact with DNA-bound TFs. They facilitate and coordinate the actions of TFs, often by bridging them to the transcription machinery or by altering the local chromatin structure. For example, the histone acetyltransferase (HAT) co-activators and the histone deacetylase (HDAC) co-repressors can increase and decrease the accessibility of DNA to TFs and GTFs, respectively, by altering the acetylation status of the histone tails [3–6]. The Swi–Snf chromatin remodeling complexes have positive roles in transcription by facilitating TFs and GTFs to access nucleosomal DNA [4,7–9].

Proteins from all three groups work together to ensure proper transcription levels of the genes inside a cell. For a given gene, the outcomes of the actions of these proteins are rather simple, with its expression level either increased (activated) or decreased (repressed). However, many TFs can work as both activators and repressors depending on cellular or promoter contexts. In addition, co-factors that are generally viewed as co-activators (e.g. the Swi–Snf complexes) or co-repressors (e.g. HDACs) can have roles that contradict their stereotypic designations. Furthermore, some GTFs have been found to be associated with co-repressor proteins to mediate gene silencing, rather than mediating transcriptional activation. This article discusses the dynamic nature of transcription control by focusing specifically on factors that participate in two opposite courses of regulation: activation and repression (Figure 1). Examples of TFs that have dual activator-repressor functions will be discussed, followed by cases of co-factors and GTFs that are engaged in distinct interactions that lead to activation and repression.

Activator-repressor switches depending on promoter and cellular contexts

Although there are bacterial TFs that can work both as activators and repressors depending on the context [10], this review will be limited to eukaryotic proteins. One well-known mammalian protein is Yin Yang 1 (YY1), a ubiquitously expressed zinc-finger TF [11,12]. YY1 has a regulatory role for many target genes, acting either positively or negatively depending on the promoter context and availability of other proteins. When the YY1-binding site overlaps with an activator-binding site, it can act as a repressor by competing with activator binding. YY1 can also bind to INR of many genes contributing positively to transcription. *In vitro* transcription experiments have shown that YY1, transcription factor IIB (TFIIB, a GTF) and RNAP are sufficient to support basal transcription from a TATA-less promoter [13]. The availability of other factors such as the adenovirus E1A protein can also affect the regulatory

Corresponding author: Jun Ma (jun.ma@cchmc.org).
Available online 19 November 2004

 0168-9525/$ - see front matter doi:10.1016/j.tig.2004.11.004

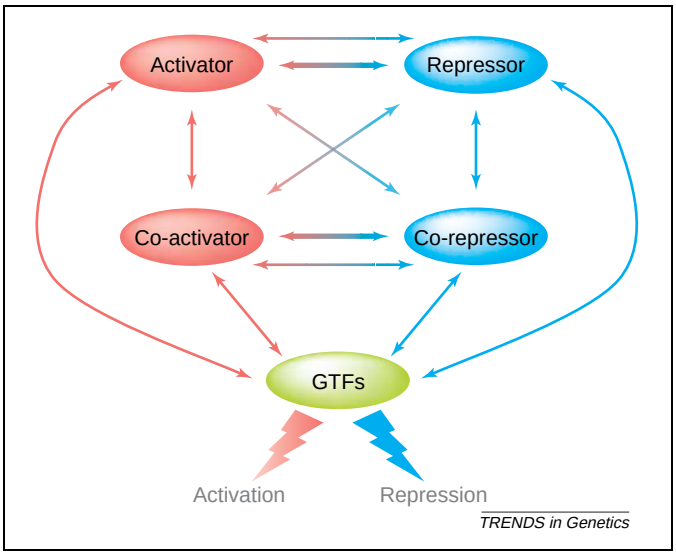

Figure 1. Functional switches and interactions in transcription control. Three broad groups of proteins – and their interactions – are involved in transcription regulation: transcription factors (activators and repressors), co-factors (co-activators and co-repressors) and general transcription factors (GTFs). The two sides of the diagram represent the two opposite sides of the regulation process: activation (left; red exemplifies the firing of the transcription machinery) and repression (right). Wide arrows indicate the functional switches of factors that cross the activation-repression line. Thin arrows represent possible functional interactions, depicting not only those in activation (red) or repression (blue) but also those that cross the activation-repression divide (two-coloured arrows). The interactions among the proteins within the same groups are not shown (e.g. activator–activator interactions, repressor–repressor interactions).

functions of YY1. It is thought that E1A can convert YY1 from a repressor to an activator by exposing a concealed activating function of YY1. Furthermore, YY1 can interact with co-factors such as the HAT co-activators [e.g. the highly related proteins p300 and CREB-binding protein (CBP)], HDAC co-repressors and a histone methyltransferase [14]. These interactions can further influence whether YY1 activates or represses transcription. A recent study demonstrates that YY1 can functionally compensate for the loss of the *Drosophila melanogaster* Polycomb group (PcG) protein Pleiohomeotic in mutant flies [15]. Because PcG proteins exist and function in large co-repressor complexes (discussed in the following section), this finding further highlights the roles of co-factors in executing the regulatory functions of YY1 *in vivo*. Interestingly, a new study suggests that YY1 is a negative regulator of p53 (discussed in the next section); however, such regulation appears to be independent of the transcriptional activity of YY1 [16].

An essential function of p53 – a tumour suppressor protein that can bind to DNA – is to activate genes involved in such essential processes as cell-cycle control, apoptosis and angiogenesis [17,18]. The following genes have been shown to be direct targets of p53: *p21* (an inhibitor of cyclin-dependent kinases), *Bax* (a pro-apoptotic factor) and thrombospondin-1 (*TSP-1*, an inhibitor of angiogenesis). It also represses many other genes through mechanisms that can be either dependent or independent of p53-binding sites. Similar to YY1, p53 can act as a repressor by competing with activators for DNA binding or by blocking the functions of DNA-bound activators. It can also repress transcription by interacting with GTFs or by recruiting co-repressors such the Sin3A–HDAC complex. It is currently not fully understood how p53 chooses to activate or repress transcription but promoter contexts, such as the p53-binding site characteristics [19,20] and the relative locations of the binding sites for p53 and other factors [21], are likely to influence the decision-making process.

Although there is a long list of TFs that possess dual activator-repressor functions, two *Drosophila* proteins are noteworthy because their functions, like many other proteins in early embryogenesis, are dependent on their concentrations. For example, the Hunchback protein (Hb) positively auto-regulates its own expression and activates *even-skipped* (*eve*) stripe 2 expression in early developing embryos [22,23]. However, Hb represses the expression of other target genes such as *knirps* (*kni*) and several other *eve* stripes in regions of the embryos where Hb is at lower concentrations [24]. What makes Hb work as an activator or a repressor in different contexts is currently not well understood but enhancer structure, Hb concentration and interactions with other proteins including GTFs and co-factors [25,26] are likely to influence such decisions. Another well-documented example is the zinc-finger protein Kruppel (Kr), which can activate or repress transcription from the same DNA-binding sites depending on its concentration [27]. At low concentrations Kr works as an activator, whereas at high concentrations it works as a repressor. It has been suggested that the monomeric and dimeric forms of Kr can interact with the GTFs, TFIIB and transcription factor IIE (TFIIE), respectively, either to activate or repress transcription [28].

Signal-dependent switches between activators and repressors

Many TFs can have distinctive activator or repressor roles in a signal-dependent manner. Members of the nuclear receptor superfamily represent excellent examples for signal-induced functional switches [29–32]. These are zinc-finger TFs that respond to ligands, for example, retinoids, steroids and thyroid hormones; some members of this family are called orphan receptors because their ligands are unknown. Nuclear receptor superfamily members have important roles in many biological processes such as differentiation, proliferation and cell death. When they are present with their ligands in a complex, these proteins activate transcription by recruiting the HAT co-activators [e.g. p300–CBP and p300–CBP-associated factor (PCAF)] and chromatin remodeling complexes to target gene promoters. However, in the absence of ligands or in the presence of antagonists, many of these proteins function as repressors by recruiting the Sin3A–HDAC complex through the nuclear receptor co-repressor (NCoR) and silencing mediator for retinoid and thyroid receptors (SMRT). Structural studies have revealed that ligands and antagonists can result in distinct conformations of the receptor proteins, thus, affecting their abilities to interact with co-activators or co-repressors [33–35].

Several other signaling pathways also involve ligand-induced functional switches of TFs [36–40]. For example, the canonical Wnt signaling pathway is dependent on the

high mobility group (HMG) proteins T-cell factor (TCF) or lymphocyte enhancer-binding factor (LEF). In the absence of Wnt signaling, TCF binds to the Wnt-response elements and represses transcription by recruiting co-repressor proteins such as Groucho (Gro) and C-terminal-binding protein (CtBP). Wnt signaling leads to the nuclear accumulation of the co-activator β-catenin, which interacts with DNA-bound TCF, recruits co-activators such as CBP and activates transcription. Similarly, in the absence of Notch signaling, the DNA-binding protein *Drosophila* Suppressor of Hairless or mammalian C-promoter-binding factor 1 [(Su(H)-CBF-1] represses transcription by interacting with co-repressors such as CtBP and Gro in *Drosophila* and the NcoR–SMRT complexes in mammals. Ligand-induced activation of the membrane-bound Notch protein leads to the release of the intracellular domain of Notch (N^{IC}), which then enters the nucleus, interacts with DNA-bound Su(H)-CBF-1, recruits co-activators such as CBP and PCAF and activates transcription. In both cases, the DNA-binding TFs [LEF–TCF in Wnt signaling and Su(H)-CBF-1 in Notch signaling] switch from repressors to activators in a ligand-dependent manner (Figure 2). In Hedgehog (Hh) signaling, the zinc-finger TF Cubitus interruptus (Ci) can enter the nucleus either as a truncated protein to repress transcription (in the absence of Hh) or as a full-length protein to activate transcription (in the presence of Hh). (For a discussion of TFs in other signaling pathways, see Barolo and Posakony [36].)

Regulation of activator functions depending on enhancer structures

Many transcriptional activators have been shown to possess domains that can inhibit their own ability to activate transcription (see examples cited in Ref. [41]). The actions of these self-inhibitory domains can range from cytoplasmic sequestration, DNA-binding inhibition to activating surface concealment. Another mechanism that regulates the function of an activator involves interactions with co-repressors. As discussed earlier, in extreme cases such interactions can completely switch the function of a TF from an activator to a repressor. The *Drosophila* morphogenetic protein Bicoid (Bcd), a homeodomain-containing TF, instructs embryonic patterning by stimulating the expression of specific genes in a concentration-dependent manner [42–44]. Although Bcd is not known to undergo activator-repressor switches, the regulation of its activity helps to illustrate the importance of the enhancer structure in selectively presenting the surfaces of a TF to enable interaction with other proteins. Recent experiments suggest that an N-terminal-located self-inhibitory domain of Bcd can interact with the co-repressor Sin3A [41,45]; Bcd can also interact with SAP18, a component of the Sin3A–HDAC complex [46]. Interestingly, the N-terminal domain of Bcd also has a crucial role in cooperative DNA binding to certain enhancers [47]. It is proposed that, depending on the arrangements of Bcd-binding sites in an enhancer, this dual-purpose N-terminal domain of Bcd is preferentially used for either cooperative DNA binding or self-inhibition [47]. This can facilitate the protein to exert its regulatory activities in a manner that is dependent not only on its own concentration but also on the enhancer structure. Recent studies further suggest that the co-activator CBP can interact with Bcd and increase its activity also in a concentration- and enhancer-dependent manner [48].

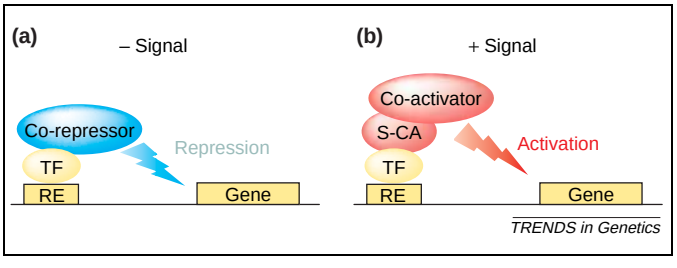

Figure 2. Signal-induced functional switches. A schematic representation of Wnt and Notch signaling events in the nucleus is shown. **(a)** In the absence of the signal, the transcription factor (TF), lymphocyte enhancer-binding factor (LEF) or T-cell factor (TCF) in Wnt signaling and *Drosophila* Suppressor of Hairless or mammalian C-promoter-binding factor 1 [Su(H)-CBF-1] in Notch signaling, binds to the signal response element (RE), recruits co-repressors and represses transcription. (b) In the presence of the signal, the signal-induced co-activator (S-CA), β-catenin in Wnt signaling and N^{IC} in Notch signaling, enters the nucleus, interacts with the TF, recruits additional co-activators and activates transcription.

Co-factors with dual roles in transcription

Historically, co-factors are loosely divided into co-activators and co-repressors (Figure 1) but their distinctions are becoming blurred. For example, CBP and p300 are generally viewed as co-activators [49,50] but in some cases they can also function as co-repressors [21,51–53]. Genome-wide microarray studies have further revealed that co-activators and co-repressors can often have roles – in some cases direct roles – that are opposite to their designations. For example, the Swi–Snf chromatin remodeling complexes are generally considered to have positive roles in transcription [4,7–9]. Indeed, microarray data in yeast show that mutations affecting components in the yeast complex reduce the transcription of many genes [54,55]. However, many other genes are expressed at increased levels in *swi–snf* mutant cells, suggesting that the Swi–Snf complex can also have negative roles in transcription. Martens and Winston studied *SER3* (a serine biosynthesis gene), one of the genes repressed by Swi–Snf [56]. They showed that the Swi–Snf complex is directly recruited to the *SER3* promoter in yeast. Interestingly, unlike transcriptional activation, which is dependent on most of the Swi–Snf subunits, only Snf2 is required for *SER3* repression. Precisely how Snf2 represses transcription is unknown, but it might involve interactions with co-repressors. It has been shown biochemically that ATP-dependent chromatin remodeling proteins can form complexes with co-repressors such as HDACs [57–60]. Interestingly, *SER3* transcription is also affected by the transcription of a non-coding gene (*SER3* regulatory gene 1 or *SRG1*) located upstream of *SER3*, but in a manner that appears to be independent of Snf2 [61].

Sin3 and HDAC proteins are generally viewed as co-repressors because they restrict the DNA accessibility to TFs and GTFs [3–6]. However, microarray experiments show that, although the expression of many genes is upregulated in yeast cells that contain either a SWI-independent 3 (*sin3*) or a *rpd3* mutant (which encodes a yeast class I HDAC), several genes are actually

expressed at reduced levels [62,63]. These results suggest that *Sin3* and *Rpd3* can also have positive roles in transcription. Kinetic experiments using the HDAC inhibitor trichostatin A (TSA) further provide evidence of a direct role of Rpd3 in the transcriptional activation of some genes [62]. In a study by Wang *et al.* [64], it was shown that another class I HDAC in yeast, encoded by *Hos2*, is required specifically for transcriptional activation. Hos2 is preferentially associated with actively transcribed genes in the genome. Interestingly, unlike Rpd3, which can deacetylate most lysines in the tails of all four histones, Hos2 exhibits a strong preference for lysines in histones H3 and H4. These results show that all histone acetylation events are not equal in their roles in regulating transcription. In addition to acetylation, histones are also subject to other types of modifications including methylation, phosphorylation and ubiquitination, and all of these modifications can work in combination to achieve distinctive regulatory outcomes [65,66].

A 'transcriptional clock' containing co-repressor proteins

In an elegant study reported recently, Metivier *et al.* [67] used the chromatin immunoprecipitation (ChIP) technique to determine the occupancy of a wide panel of GTFs and co-factors at an oestrogen-inducible promoter. It was revealed that these factors join and depart from the transcription complexes formed at the promoter in a cyclic manner. This sequential and combinatorial assembly of transcription complexes at the promoter was referred to as a 'transcriptional clock', with each cycle lasting ~40 minutes. Surprisingly, co-repressor proteins HDAC1 and HDAC7 are also integral parts of this transcriptional clock. The loading and exiting of these proteins, together with the Swi–Snf chromatin-remodeling complex, appear to mark the end of each cycle and prepare the promoter for the next cycle. Although it is currently unknown whether transcriptional activation at other promoters requires the same type of waves of complex formation, the findings described in this study clearly illustrate that proteins that are generally viewed as co-repressors (such as HDACs) actually have integral roles during the transcription-activation process.

General transcription factors mediating gene silencing

According to a well-established recruitment model, transcriptional activation is a process of bringing, ultimately, the transcription machinery to a promoter [68,69]. However, recent studies in yeast and *Drosophila* suggest that gene silencing can act following the assembly of the transcription machinery. Cell-type specification in the bakers yeast *Saccharomyces cerevisiae* is controlled by gene products encoded by the *MAT* locus. In addition to this active locus, there are two donor cassettes (*HMR* and *HML*) that are silenced by *cis*-elements called silencers. Several proteins, including repressor activator protein 1 (RAP1), ARS-binding factor 1 (ABF1) and origin replication complex (ORC), bind to DNA sequences in the silencers and recruit the silencing information regulator (SIR) complex containing Sir2, Sir3 and Sir4 [70,71]. Sir2 is a nicotinamide adenine dinucleotide (NAD)-dependent HDAC (class III), whereas Sir3 and Sir4 (which are not HDACs) can interact with the hypoacetylated N-terminal tails of histones H3 and H4 to enable the complex to propagate along nucleosomes. Because silenced chromatin is refractory to sequence-specific DNA-binding proteins such as restriction enzymes [72], it was thought, according to one model, that SIR-dependent hypoacetylation of chromatin at the silent cassettes prevented activators and GTFs from accessing DNA. However, a recent report argued against this simple idea [73]. Using ChIP assays, Sekinger and Gross showed that the SIR-silenced *Hsp82* promoter is actually occupied by not only the activator heat shock factor (HSF) but also components of the transcription machinery such as TBP and RNAP. Both TBP and RNAP are also present at the promoter of *HMRa1*, a natural target gene silenced by the SIR complex. Precisely how the assembled transcription machinery remains inactive at the SIR-silenced genes is not well understood but, as discussed in the next section, one possible mechanism might involve direct interactions between co-repressors and GTFs.

Studies of homeotic gene regulation in *Drosophila* have revealed specific interactions between co-repressors and GTFs [74,75]. Homeotic genes themselves encode TFs that control segment identity. In early embryogenesis, the expression of these genes is initiated by DNA-binding TFs such as those encoded by gap and pair-rule genes. The expression profiles of the homeotic genes are subsequently controlled by the *trithorax* group (*trxG*) and *PcG* genes, which maintain their active and silenced states, respectively (Figure 3). TrxG proteins can form chromatin remodeling complexes related to the yeast Swi–Snf complex [76–78], whereas PcG proteins assemble into large complexes that can contain co-repressors such as Sin3A and HDAC1 [79,80]. Paradoxically, PcG and trxG proteins often co-localize on DNA, suggesting that these proteins with opposite functions can work in concert to regulate transcription [81,82]. More intriguingly, a PcG repressive complex (PRC1) actually contains several GTFs such as TBP and TBP-associated factors (TAFs) [80]. In addition, ChIP experiments show that many PcG-silenced promoters are occupied by components of the transcription machinery including TBP, TFIIB and RNAP [83,84]. Together, these studies of both SIR-mediated silencing in yeast and PcG-mediated repression in *Drosophila* suggest that one mechanism of gene silencing is to keep the pre-assembled transcription machinery inactive, probably through interactions between GTFs and co-repressors (Figure 3). This repression mechanism is distinct from the mechanism where repressors interact with GTFs to prevent the assembly of the transcription machinery [28,85].

Three words important in transcription control: context, context, context

It is evident that regulation of gene expression is a highly dynamic process requiring the actions of – and communications between – many proteins. Although these proteins help make unambiguous and simple decisions to either activate or repress a gene, their roles and interactions often cross the activation-repression line. Whether a TF

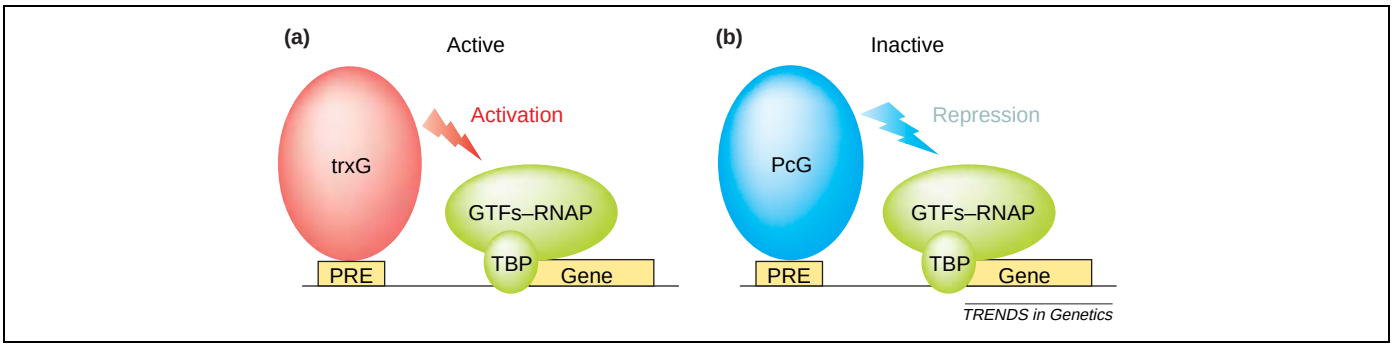

Figure 3. Homeotic gene expression. A schematic representation depicts the roles of trithorax group (trxG) co-activators and Polycomb group (PcG) co-repressors in maintaining **(a)** the active and **(b)** inactive expression states of a homeotic gene, respectively. Note that in both cases that the general transcription factors (GTFs), including TATA-binding protein (TBP) and RNA polymerase (RNAP), are present at the promoter. Both trxG and PcG proteins can exert their effects through Polycomb response elements (PREs), by interacting with proteins (not shown) that recognize specific DNA sequences in these elements or by recognizing epigenetic marks (e.g. methylated histones) at these locations [74,75].

activates or represses transcription of a given gene is determined by the specific microenvironment in which it operates, including such parameters as: (i) its own concentration and physical form (e.g. ligand interaction or modifications including ubiquitination and SUMOylation [86]); (ii) how it interacts with other TFs on DNA (e.g. competitive or cooperative DNA binding); and (iii) the types of surfaces (e.g. activation versus repression) that are available for interacting with co-factors and/or GTFs upon 'landing' on an enhancer. The availability and relative concentrations of specific co-activators and co-repressors, and the chromatin architecture and initial expression status of a gene are also important variables that can influence the ability of a TF to activate or repress a gene. Similarly, the ability of a co-factor to activate or repress transcription is also influenced by the micro-environment in which it works and, as discussed previously, its role can be time-dependent. Even GTFs can interact with co-repressors to mediate gene silencing, in addition to their roles of interacting with activators and co-activators to mediate gene activation. It is clear that proteins that control transcription have multiple 'personalities': their roles depend on when and where they are in the transcription process and on the proteins that are around them.

Acknowledgements

I thank members of my laboratory and M. Barton, K. Campbell, T. Cook, I. Cartwright, R. Hegde and X. Lin for discussions or comments of the manuscript. Owing to space constraints some articles could not be discussed. This work is supported in part by grants from the AHA, NSF and DOD.

References

1 Kadonaga, J.T. (2004) Regulation of RNA polymerase II transcription by sequence-specific DNA binding factors. *Cell* 116, 247–257
2 Levine, M. and Tjian, R. (2003) Transcription regulation and animal diversity. *Nature* 424, 147–151
3 Yang, X.J. and Seto, E. (2003) Collaborative spirit of histone deacetylases in regulating chromatin structure and gene expression. *Curr. Opin. Genet. Dev.* 13, 143–153
4 Narlikar, G.J. *et al*. (2002) Cooperation between complexes that regulate chromatin structure and transcription. *Cell* 108, 475–487
5 Kurdistani, S.K. and Grunstein, M. (2003) Histone acetylation and deacetylation in yeast. *Nat. Rev. Mol. Cell Biol.* 4, 276–284
6 Peterson, C.L. (2002) HDAC's at work: everyone doing their part. *Mol. Cell* 9, 921–922
7 Martens, J.A. and Winston, F. (2003) Recent advances in understanding chromatin remodeling by Swi/Snf complexes. *Curr. Opin. Genet. Dev.* 13, 136–142
8 Berger, S.L. and Felsenfeld, G. (2001) Chromatin goes global. *Mol. Cell* 8, 263–268
9 Peterson, C.L. and Workman, J.L. (2000) Promoter targeting and chromatin remodeling by the SWI/SNF complex. *Curr. Opin. Genet. Dev.* 10, 187–192
10 Ptashne, M. (1986) *A genetic switch*, Cell and Blackwell scientific publications
11 Shi, Y. *et al*. (1997) Everything you have ever wanted to know about Yin Yang 1. *Biochim. Biophys. Acta* 1332, 49–66
12 Thomas, M.J. and Seto, E. (1999) Unlocking the mechanisms of transcription factor YY1: are chromatin modifying enzymes the key? *Gene* 236, 197–208
13 Usheva, A. and Shenk, T. (1994) TATA-binding protein-independent initiation: YY1, TFIIB, and RNA polymerase II direct basal transcription on supercoiled template DNA. *Cell* 76, 1115–1121
14 Rezai-Zadeh, N. *et al*. (2003) Targeted recruitment of a histone H4-specific methyltransferase by the transcription factor YY1. *Genes Dev.* 17, 1019–1029
15 Atchison, L. *et al*. (2003) Transcription factor YY1 functions as a PcG protein *in vivo*. *EMBO J.* 22, 1347–1358
16 Sui, G. *et al*. (2004) Yin Yang 1 is a negative regulator of p53. *Cell* 117, 859–872
17 Ho, J. and Benchimol, S. (2003) Transcriptional repression mediated by the p53 tumour suppressor. *Cell Death Differ.* 10, 404–408
18 Vousden, K.H. and Lu, X. (2002) Live or let die: the cell's response to p53. *Nat. Rev. Cancer* 2, 594–604
19 Hoffman, W.H. *et al*. (2002) Transcriptional repression of the anti-apoptotic survivin gene by wild type p53. *J. Biol. Chem.* 277, 3247–3257
20 Johnson, R.A. *et al*. (2001) Transcriptional repression by p53 through direct binding to a novel DNA element. *J. Biol. Chem.* 276, 27716–27720
21 Lee, K.C. *et al*. (1999) p53-mediated repression of alpha-fetoprotein gene expression by specific DNA binding. *Mol. Cell. Biol.* 19, 1279–1288
22 Simpson-Brose, M. *et al*. (1994) Synergy between the hunchback and bicoid morphogens is required for anterior patterning in *Drosophila*. *Cell* 78, 855–865
23 Small, S. *et al*. (1991) Transcriptional regulation of a pair-rule stripe in *Drosophila*. *Genes Dev.* 5, 827–839
24 Clyde, D.E. *et al*. (2003) A self-organizing system of repressor gradients establishes segmental complexity in Drosophila. *Nature* 426, 849–853
25 Sauer, F. *et al*. (1995) Multiple TAFIIs directing synergistic activation of transcription. *Science* 270, 1783–1788
26 Kehle, J. *et al*. (1998) dMi-2, a hunchback-interacting protein that functions in polycomb repression. *Science* 282, 1897–1900
27 Sauer, F. and Jackle, H. (1991) Concentration-dependent transcriptional activation or repression by kruppel from a single binding site. *Nature* 353, 563–566
28 Sauer, F. *et al*. (1995) Control of transcription by kruppel through interactions with TFIIB and TFIIEβ. *Nature* 375, 162–164
29 Xu, L. *et al*. (1999) Coactivator and corepressor complexes in nuclear receptor function. *Curr. Opin. Genet. Dev.* 9, 140–147

30 Zhang, J. and Lazar, M.A. (2000) The mechanism of action of thyroid hormones. *Annu. Rev. Physiol.* 62, 439–466
31 Aranda, A. and Pascual, A. (2001) Nuclear hormone receptors and gene expression. *Physiol. Rev.* 81, 1269–1304
32 Park, J.I. *et al*. (2003) Molecular mechanism of chicken ovalbumin upstream promoter-transcription factor (COUP-TF) actions. *Keio J. Med.* 52, 174–181
33 Moras, D. and Gronemeyer, H. (1998) The nuclear receptor ligand-binding domain: structure and function. *Curr. Opin. Cell Biol.* 10, 384–391
34 Renaud, J.P. and Moras, D. (2000) Structural studies on nuclear receptors. *Cell. Mol. Life Sci.* 57, 1748–1769
35 Nettles, K.W. and Greene, G.L. (2003) Nuclear receptor ligands and cofactor recruitment: is there a coactivator 'on deck'? *Mol. Cell* 11, 850–851
36 Barolo, S. and Posakony, J.W. (2002) Three habits of highly effective signaling pathways: principles of transcriptional control by developmental cell signaling. *Genes Dev.* 16, 1167–1181
37 Bienz, M. (1998) TCF: transcriptional activator or repressor? *Curr. Opin. Cell Biol.* 10, 366–372
38 Kalderon, D. (2002) Similarities between the hedgehog and Wnt signaling pathways. *Trends Cell Biol.* 12, 523–531
39 Aza-Blanc, P. and Kornberg, T.B. (1999) Ci: a complex transducer of the hedgehog signal. *Trends Genet.* 15, 458–462
40 Lai, E.C. (2002) Keeping a good pathway down: transcriptional repression of Notch pathway target genes by CSL proteins. *EMBO Rep.* 3, 840–845
41 Zhao, C. *et al*. (2002) The activity of the *Drosophila* morphogenetic protein bicoid is inhibited by a domain located outside its homeodomain. *Development* 129, 1669–1680
42 Driever, W. (1992) The bicoid morphogen: concentration dependent transcriptional activation of zygotic target genes during early *Drosophila* development. In *Transcriptional Regulation* (Yamamoto, K. ed.), pp. 1221–1250, Cold Spring Harbor Laboratory Press
43 Rivera-Pomar, R. and Jackle, H. (1996) From gradients to stripes in *Drosophila* embryogenesis: filling in the gaps. *Trends Genet.* 12, 478–483
44 Ephrussi, A. and Johnston, D.S. (2004) Seeing is believing. The bicoid morphogen gradient matures. *Cell* 116, 143–152
45 Zhao, C. *et al*. (2003) A composite motif of the *Drosophila* morphogenetic protein bicoid critical to transcription control. *J. Biol. Chem.* 278, 43901–43909
46 Zhu, W. *et al*. (2001) *Drosophila* SAP18, a member of the Sin3/Rpd3 histone deacetylase complex, interacts with Bicoid and inhibits its activity. *Dev. Genes Evol.* 211, 109–117
47 Fu, D. *et al*. (2003) Enhancer sequences influence the role of the amino terminal domain of bicoid in transcription. *Mol. Cell. Biol.* 23, 4439–4448
48 Fu, D. *et al*. The co-activator CBP participates in enhancer-dependent activities of bicoid. *J. Biol. Chem.* (in press)
49 Goodman, R.H. and Smolik, S. (2000) CBP/p300 in cell growth, transformation, and development. *Genes Dev.* 14, 1553–1577
50 Chan, H.M. and La Thangue, N.B. (2001) p300/CBP proteins: HATs for transcriptional bridges and scaffolds. *J. Cell Sci.* 114, 2363–2373
51 Waltzer, L. and Bienz, M. (1998) *Drosophila* CBP represses the transcription factor TCF to antagonize Wingless signalling. *Nature* 395, 521–525
52 Jayachandra, S. *et al*. (1999) Three unrelated viral transforming proteins (vIRF, EBNA2, and E1A) induce the MYC oncogene through the interferon-responsive PRF element by using different transcription coadaptors. *Proc. Natl. Acad. Sci. U. S. A.* 96, 11566–11571
53 Wolf, D. *et al*. (2002) Acetylation of beta-catenin by CREB-binding protein (CBP). *J. Biol. Chem.* 277, 25562–25567
54 Sudarsanam, P. *et al*. (2000) Whole-genome expression analysis of snf/swi mutants of *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U. S. A.* 97, 3364–3369
55 Holstege, F.C. *et al*. (1998) Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* 95, 717–728
56 Martens, J.A. and Winston, F. (2002) Evidence that Swi/Snf directly represses transcription in *S. cerevisiae*. *Genes Dev.* 16, 2231–2236
57 Zhang, Y. *et al*. (1998) The dermatomyositis-specific autoantigen Mi2 is a component of a complex containing histone deacetylase and nucleosome remodeling activities. *Cell* 95, 279–289
58 Xue, Y. *et al*. (1998) NURD, a novel complex with both ATP-dependent chromatin-remodeling and histone deacetylase activities. *Mol. Cell* 2, 851–861
59 Tong, J.K. *et al*. (1998) Chromatin deacetylation by an ATP-dependent nucleosome remodelling complex. *Nature* 395, 917–921
60 Sif, S. *et al*. (2001) Purification and characterization of mSin3A-containing Brg1 and hBrm chromatin remodeling complexes. *Genes Dev.* 15, 603–618
61 Martens, J.A. *et al*. (2004) Intergenic transcription is required to repress the *Saccharomyces cerevisiae* SER3 gene. *Nature* 429, 571–574
62 Bernstein, B.E. *et al*. (2000) Genome wide studies of histone deacetylase function in yeast. *Proc. Natl. Acad. Sci. U. S. A.* 97, 13708–13713
63 Fazzio, T.G. *et al*. (2001) Widespread collaboration of Isw2 and Sin3–Rpd3 chromatin remodeling complexes in transcriptional repression. *Mol. Cell. Biol.* 21, 6450–6460
64 Wang, A. *et al*. (2002) Requirement of Hos2 histone deacetylase for gene activity in yeast. *Science* 298, 1412–1414
65 Jenuwein, T. and Allis, C.D. (2001) Translating the histone code. *Science* 293, 1074–1080
66 Wu, J. and Grunstein, M. (2000) 25 years after the nucleosome model: chromatin modifications. *Trends Biochem. Sci.* 25, 619–623
67 Metivier, R. *et al*. (2003) Estrogen receptor-alpha directs ordered, cyclical, and combinatorial recruitment of cofactors on a natural target promoter. *Cell* 115, 751–763
68 Ptashne, M. and Gann, A. (1997) Transcriptional activation by recruitment. *Nature* 386, 569–577
69 Stargell, L.A. and Struhl, K. (1996) Mechanisms of transcriptional activation *in vivo*: two steps forward. *Trends Genet.* 12, 311–315
70 Rusche, L.N. *et al*. (2003) The establishment, inheritance, and function of silenced chromatin in *Saccharomyces cerevisiae*. *Annu. Rev. Biochem.* 72, 481–516
71 Blander, G. and Guarente, L. (2004) The Sir2 family of protein deacetylases. *Annu. Rev. Biochem.* 73, 417–435
72 Loo, S. and Rine, J. (1994) Silencers and domains of generalized repression. *Science* 264, 1768–1771
73 Sekinger, E.A. and Gross, D.S. (2001) Silenced chromatin is permissive to activator binding and PIC recruitment. *Cell* 105, 403–414
74 Orlando, V. (2003) Polycomb, epigenomes, and control of cell identity. *Cell* 112, 599–606
75 Levine, S.S. *et al*. (2004) Division of labor in polycomb group repression. *Trends Biochem. Sci.* 29, 478–485
76 Collins, R.T. *et al*. (1999) Osa associates with the brahma chromatin-remodeling complex and promotes the activation of some target genes. *EMBO J.* 18, 7029–7040
77 Papoulas, O. *et al*. (1998) The *Drosophila* trithorax group proteins BRM, ASH1 and ASH2 are subunits of distinct protein complexes. *Development* 125, 3955–3966
78 Kal, A.J. *et al*. (2000) The *Drosophila* brahma complex is an essential coactivator for the trithorax group protein zeste. *Genes Dev.* 14, 1058–1071
79 Tie, F. *et al*. (2001) The *Drosophila* polycomb group proteins ESC and E(Z) are present in a complex containing the histone-binding protein p55 and the histone deacetylase RPD3. *Development* 128, 275–286
80 Saurin, A.J. *et al*. (2001) A *Drosophila* polycomb group complex includes Zeste and dTAFII proteins. *Nature* 412, 655–660
81 Chinwalla, V. *et al*. (1995) The *Drosophila* trithorax protein binds to specific chromosomal sites and is co-localized with Polycomb at many sites. *EMBO J.* 14, 2056–2065
82 Orlando, V. *et al*. (1998) Binding of trithorax and polycomb proteins to the bithorax complex: dynamic changes during early *Drosophila* embryogenesis. *EMBO J.* 17, 5141–5150
83 Breiling, A. *et al*. (2001) General transcription factors bind promoters repressed by polycomb group proteins. *Nature* 412, 651–655
84 Dellino, G.I. *et al*. (2004) Polycomb silencing blocks transcription initiation. *Mol. Cell* 13, 887–893
85 Li, C. and Manley, J.L. (1998) Even-skipped represses transcription by binding TATA binding protein and blocking the TFIID–TATA box interaction. *Mol. Cell. Biol.* 18, 3771–3781
86 Gill, G. (2004) SUMO and ubiquitin in the nucleus: different functions, similar mechanisms? *Genes Dev.* 18, 2046–2059